# Berry-Esseen bounds for arbitrary statistics

een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen in het
openbaar te verdedigen op 26 januari 2000
des namiddags om 1.30 uur precies
door

**Ivo Benedikt Alberink**

geboren op 5 juni 1971
te Hengelo (ov)

**Promotor:**
Prof. Dr. W.Th.F. den Hollander

**Co-promotor:**
Dr. M.C.A. van Zuijlen

**Manuscriptcommissie:**
Prof. Dr. F. Götze (University of Bielefeld, Germany)
Dr. R. Helmers (Centrum voor Wiskunde en Informatica, Amsterdam)
Prof. Dr. G. Pap (Lajos Kossuth University, Debrecen, Hungary)

# Contents

# Chapter 1

# Introduction

Let $X_1, \ldots, X_n$ be independent, not necessarily identically distributed random variables, taking their values in arbitrary measurable spaces $(\mathcal{X}_j, \mathcal{B}_j)$, and consider arbitrary real-valued statistics

$$\mathbb{T} \;=\; \mathbb{T}(n) \;=\; \mathbb{T}(X_1, \ldots, X_n)$$

with existing first moments. A sequence $\mathbb{T} = \mathbb{T}(n)$ of statistics is called *asymptotically normal* if there exist normalizing constants $a_n, b_n \in \mathbb{R}$ such that the sequence $(\mathbb{T} - a_n)/b_n$ converges weakly to the standard normal distribution. Speaking very generally, our main objective is to develop a sufficiently general and exhausting mathematical method to analyze asymptotically normal statistics, without making further assumptions on their structure. In fact, we are interested in obtaining estimates of the accuracy of normal approximations, and, for example, bounds for the remainder terms in asymptotic (Edgeworth) expansions of $\mathbb{T}$'s distribution. Under various assumptions on the structure of $\mathbb{T}$, a large number of results of this kind is already available. Since this number is increasing steadily, it seems worthwhile to have a unifying approach.

In order to tackle the afore-mentioned problem in a way which does not presuppose that $\mathbb{T}$ has a certain structure, the classical approach is to write $\mathbb{T}$ by its canonical Hoeffding decomposition

$$\mathbb{T} \;=\; \textstyle\sum_{A \subset \{1, \ldots, n\}} T_A$$

and look for results in terms of moments of the $T_A$. In Appendix A we describe Hoeffding's decomposition and the related concept of differences of $\mathbb{T}$. The two are of great importance in the formulation of both our results and their proofs, whereas differences have as well nice properties from a practical point of view. Elementary properties of the two are gathered in the appendix and used throughout the manuscript, so the reader is strongly advised to take a look at it before proceeding to the Chapters 2, 3 and 4.

As an example of a rather general class of statistics we might look at the class of symmetric statistics based on independent, identically distributed (i.i.d.) samples. By *symmetric* we mean that the statistics are invariant under permutations of their arguments. This class has been studied, e.g., by Dynkin and Mandelbaum (1983), Mandelbaum and Taqqu (1984), Giraitis and Surgailis (1984) and Mandrekar and Meerschaert (1994), and normal approximations have been considered by Van Zwet (1984), Friedrich (1989), Lai and Wang (1993), Guan (1994), Bentkus, Götze and Van Zwet (1997), Putter and Van Zwet (1998) and Wang (1998). Being nice and convenient from a mathematical point of view, the i.i.d. and symmetry assumptions are rather restrictive in applications. For example, in the case of linear rank statistics, or statistics depending on two (or more) samples, the assumption of symmetry is usually not applicable, whereas in the latter case usually the i.i.d. assumption is not satisfied as well. To our knowledge, the only bounds for general statistics that have been obtained so far without using the i.i.d. and symmetry assumptions are due to Friedrich (1989). Friedrich's bounds however do not differ very much from the ones obtained in the i.i.d. and symmetric case in the sense that they are stated in terms of quantities like $\max_{1\le j\le n} \mathbb{E}\,|T_j|^3$. A more natural way of dropping the two assumptions would for example involve that the bounds for the accuracy of approximations depend in an additive way on separate observations. In this case we would be looking at quantities like $\sum_{j=1}^n \mathbb{E}\,|T_j|^3$, as in the classical non-i.i.d. bound for sums of independent random variables (see Feller (1971), or Petrov (1995)). A first step in this direction was made by Bentkus, Bloznelis and Götze (1996), who obtained additive bounds in the case of Student's statistic and self-normalized sums. The methods to obtain such results involve rather refined techniques, based mainly on non-linear Fourier analysis, though direct methods to analyze the distributions may sometimes be applied as well.

The main results are obtained on two closely related topics. The first one, considered in Chapter 2, concerns concentration bounds, and amounts to finding upper bounds for the maximum probability for $\mathbb{T}$ of being in any closed interval of fixed length $\lambda$. A review on concentration can be found, for example, in Hengartner and Theodorescu (1973). In statistics, its practical importance seems to be limited, but it is interesting from a mathematical point of view. For example, certain typical problems in analytic number theory may be stated in terms of bounds for concentrations of measures. The technique we develop to obtain the concentration bounds seems to be also useful in a more refined analysis of the distribution of $\mathbb{T}$, for example when one is looking for estimates of convergence rates to normality or asymptotic expansions. Estimates of this form are as well important for the application of modern smoothing techniques, see Prawitz (1972), Bentkus (1994), Aleshkyavichene and Statulevichius (1997), Bentkus, Götze and Van Zwet (1997). In main lines our technique consists of using a variant of Esseen's smoothing lemma, which reduces the problem to the estimation of an integral, with the characteristic function of $\mathbb{T}$ as its integrand. This integral can be handled by means of Taylor expansions and conditioning arguments that may be found in Van Zwet (1984). The non-i.i.d. result is achieved in a similar but more refined way, using in addition a randomization of the sample elements. The bounds that we obtain do not involve requirements on the structure of $\mathbb{T}$ and have a rather simple additive structure. In fact, they depend on the variance of the linear part of the statistic, the sum of the $\mathbb{E}\,|T_j|^3$, and, roughly speaking, the variance of the non-linear part. The non-i.i.d. result generalizes the result in the symmetric, identically distributed case. The bounds are precise, up to absolute constants, in the sense that applied to special classes of statistics they always provide optimal results.

The second topic, under consideration in Chapter 3, is that of Berry-Esseen bounds. These are in fact upper bounds for the maximum distance between the distribution function of a standardized version of $\mathbb{T}$ and the standard normal one. They have been investigated, among others, by Chan and Wierman (1977), Callaert and Janssen (1978), Helmers and Van Zwet (1983) and Van Zwet (1984), who obtained this type of results in increasing generality in the symmetric, i.i.d. case, and by Friedrich (1989), who considered the general, independent case. Our results are again given in terms of the variance of the linear part of $\mathbb{T}$, the summed $\mathbb{E}\,|T_j|^3$ and the variance of the non-linear part. The method to them is the same as the one in order to get the concentration bounds, now starting from Esseen's smoothing lemma. In the symmetric, i.i.d. case we get a bit better result than in the general case, which yields a slightly better result than the one in Van Zwet (1984), to our knowledge the best known result in the symmetric, i.i.d. case. In the non-i.i.d. case our result is far more transparant than, and improves the result of Friedrich (1989), since it uses no constructs like maxima over $\mathbb{E}\,|T_j|^3$, and it looks much nicer.

In the final chapter we consider applications of the obtained results. We first look at simple linear rank statistics, obtaining results that are as good as the ones obtained by Does (1982). After this we look at $U$-statistics and so-called *incomplete $U$-statistics*, starting from i.i.d. samples. For $U$-statistics we get a result that is a bit better than that of Van Zwet (1984). For incomplete $U$-statistics no Berry-Esseen bounds were established up to now, so we cannot compare our results to earlier ones, but our bounds allow us to reprove central limit theorems by Blom (1976) and Janson (1984). Finally we turn to some *self-normalized* statistics, statistics of the form

$$\mathbb{T} = R/\sqrt{S}, \tag{1.1}$$

where $R$ and $S \geq 0$ are certain statistics. In fact, taking the (trivial) normalizer $S \equiv 1$, any statistic $\mathbb{T}$ may be considered as self-normalized. The simplest non-trivial example is provided by the self-normalized sum

$$\mathbb{T} = \frac{X_1 + \ldots + X_n}{\sqrt{X_1^2 + \ldots + X_n^2}},$$

which usually serves as a good example to test the strength of Berry-Esseen bounds for general statistics. We find concentration bounds that are as good as may be expected. Contrary to this, it seems not possible to obtain general Berry-Esseen bounds which lead to optimal results for self-normalized sums. Looking at Student's statistic in a one- and two-sample setting, and at two standard statistics associated to the model of linear regression statistic, we obtain Berry-Esseen bounds that require the existence of $3 + \delta^{\text{th}}$ moments, for some fixed $\delta > 0$. In the case of the two-sample Student statistic this is a new result, though probably as well not the best possible. A way to overcome this problem would be to consider general statistics of form (1.1) and to obtain bounds which use moments of differences of both $R$ and $S$. Clearly this would not lead to loss of generality.

# Chapter 2

# Concentration functions

## 2.1 Introduction and results

First we gather some general notation. Here and throughout the manuscript we shall write $\mathbb{T}$ instead of $\mathbb{T}(n)$, and in general use notation that does not reflect dependence on $n$. Moreover, let

$$N \ := \ \{1, \dots, n\}.$$

The cardinality of any subset $A \subset N$ will be denoted by $|A|$, and its complement by $A^c$. For $A \subset N$ we use the conditional expectations $\mathbb{E}(\mathbb{T} \,|\, A)$ and $\mathbb{E}_A \mathbb{T}$ as in (A.2) and (A.3). For any real number $x$, by $[x]$ we mean its integer part. For any bounded function $f$, by

$$\|f\|_\infty \ := \ \sup_{x \in \mathbb{R}} |f(x)|$$

we denote its supremum norm. The notation

$$b_1 \ \ll \ b_2, \tag{2.1}$$

for $b_1, b_2 \geq 0$, will mean by definition that $b_1 \leq c\, b_2$ for some absolute constant $c$. If the same is true for a constant $c(p)$ depending on some parameter $p$, we will write $b_1 \ll_p b_2$. Finally, for $y \in \mathbb{R}$, we shall write

$$e\{y\} \ := \ \exp\{iy\},$$

with $i = \sqrt{-1}$ denoting the complex root.

Let $\mathbb{T} = \mathbb{T}(X_1, \dots, X_n)$ as before denote an arbitrary statistic, depending on an independent, not necessarily identically distributed sample, with $\mathbb{E}|\mathbb{T}| < \infty$, and let $\sigma^2 := \operatorname{var}\mathbb{T}$, which may be $\infty$. As in (A.6) and (A.7), we use $\mathbb{T}$'s Hoeffding decomposition

$$\mathbb{T} \;=\; \textstyle\sum_{A\subset\{1,\dots,n\}} T_A \;=\; \mathbb{E}\,\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2 + \mathbf{T}_3.$$

The *concentration function* of our statistic $\mathbb{T}$, defined by

$$Q(\mathbb{T}, \lambda) \;:=\; \sup_{x\in\mathbb{R}} \mathbb{P}(x \le \mathbb{T} \le x + \lambda), \quad \text{all } \lambda \ge 0,$$

gives a measure for the concentration of probability mass on (small) intervals. Concentration functions were studied, for example, by Hengartner and Theodorescu (1973) and Petrov (1995). In this manuscript we are interested in finding upper bounds for them. Although the practical importance of these is not very big, they are interesting from a theoretical point of view. A second reason for obtaining these bounds is that the method to get to them functions nicely as a preliminary stage in the enterprise of finding Berry-Esseen bounds for arbitrary statistics. The practical value of the latter is clear.

Specialized to the case of symmetric $\mathbb{T}$ and i.i.d. samples, taking $s^2$ as in (A.14) and $\tilde\beta_3$ and $\tilde\Delta_2^2$ as in (A.20), we have the following result:

**Theorem 2.1.** *Suppose that $0 < s^2 < \infty$, the sample $X_1, \dots, X_n$ is i.i.d. and $\mathbb{T}$ is symmetric. Then there exists an absolute constant $c$ such that, for all $\lambda \ge 0$,*

$$Q(\mathbb{T}/s, \lambda) \;\le\; c\,\max\{\lambda,\ \tilde\beta_3\, s^{-3}\, n^{-1/2},\ \tilde\Delta_2^2\, s^{-2}\, n^{-1}\}. \tag{2.2}$$

*The bound holds as well with $s$ replaced by $\sigma$ if in addition $\sigma^2 < \infty$.*

The constant $c$ in (2.2) satisfies $c \le 850$. The bound (2.2) reflects that the concentration function of $\mathbb{T}$ depends mainly on its linear part $\mathbb{T}_1$. To be more specific, if $\tilde\beta_3\, s^{-3}$ and $\tilde\Delta_2^2\, s^{-2}$ are bounded from above as $n \to \infty$, the contribution of $\mathbb{T}_1$ is of order $\mathcal{O}(n^{-1/2})$, and the contribution of all higher order terms together is $\mathcal{O}(n^{-1})$.

For our main result, we look at the general case in which the sample $X_1, \dots, X_n$ is independent but not necessarily identically distributed, and $\mathbb{T}$

is an arbitrary function, that is, not necessarily symmetric in its arguments. Let $s^2$ be as in (A.14) and $\beta$ and $\Delta^2$ as in (A.15). The result reads as follows:

**Theorem 2.2.** *Suppose that $0 < s^2 < \infty$, the observations $X_1, \dots, X_n$ are independent and $\mathbb{T}$ is arbitrary. Then there exists an absolute constant $c$ such that, for all $\lambda \geq 0$,*

$$Q(\mathbb{T}/s, \lambda) \;\leq\; c \max\left\{\lambda,\ \beta/s^3,\ \Delta^2/s^2\right\}.$$

*The bound holds as well with $s$ replaced by $\sigma$ if in addition $\sigma^2 < \infty$.*

Theorem 2.2 is a direct generalization of Theorem 2.1. Indeed, in the case of an i.i.d. sample and a symmetric statistic, using (A.24) we have that

$$\beta \;=\; \tilde{\beta}_3\, n^{-1/2} \quad \text{and} \quad \Delta^2 \;\leq\; \tfrac{1}{2}\, \tilde{\Delta}_2^2\, n^{-1}. \tag{2.3}$$

We start by giving the proof of Theorem 2.1. For reasons of transparancy, in Section 2.2 we first consider the special case of $U$-statistics of order 2, that is, the case where in the Hoeffding decomposition (A.6) the part $\mathbf{T}_3 = 0$. In Section 2.3 we then extend this proof to the case of arbitrary statistics. Using the same scheme, the proof of Theorem 2.2 will be given in the Sections 2.4 and 2.5.

In Chapter 4 we provide applications of the Theorems 2.1 and 2.2 to particular statistics, each time obtaining optimal results. Self-normalized statistics are of special interest here. In these applications the exact knowledge of Hoeffding decompositions is often not important at all, since instead of them one can use so-called differences, see (A.8). Especially looking at (A.13), we have that

$$\beta \;\leq\; \textstyle\sum_{j=1}^n \mathbb{E}\,|D_j\mathbb{T}|^3 \quad \text{and} \quad \Delta^2 \;=\; \textstyle\sum_{1\leq j<k\leq n} \mathbb{E}\,|D_j D_k \mathbb{T}|^2.$$

In the symmetric i.i.d. case this corresponds to:

$$\tilde{\beta}_3 \;\leq\; n^{3/2}\, \mathbb{E}\,|D_1\mathbb{T}|^3 \quad \text{and} \quad \tilde{\Delta}_2^2 \;=\; n^3\, \mathbb{E}\,|D_1 D_2 \mathbb{T}|^2.$$

## 2.2 Proof of Theorem 2.1 for $U$-statistics of order 2

For our bounds we will be using the following inequality, which holds for arbitrary $\mathbb{T}$ and may be found in Petrov (1995) as Lemma 1.16: for any

$a > 0$,

$$Q(\mathbb{T}, \lambda) \le 2 \cdot (\tfrac{96}{95})^2 \max\{\lambda, a^{-1}\} \int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt. \tag{2.4}$$

As promised, first we will prove Theorem 2.1 in the case where $\mathbb{T}$ is a so-called $U$-statistic of order 2, that is, $\mathbb{T} = \mathbb{E}\,\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2$ (cf. (A.6)). Taking $\tilde{\gamma}_2$ as in (A.20), in this case $\tilde{\Delta}_2^2 = \tilde{\gamma}_2$.

Without loss of generality we may assume that

$$\mathbb{E}\,\mathbb{T} = 0 \quad \text{and} \quad s^2 = 1.$$

Indeed, the expectation value of $\mathbb{T}$ may be taken zero as

$$Q(\mathbb{T}, \lambda) = Q(\mathbb{T} - \mathbb{E}\,\mathbb{T}, \lambda).$$

In case $s^2 \ne 1$ we may just apply the reduced theorem to $\mathbb{T}/s$. We prove Theorem 2.1 by using (2.4), taking

$$a := \min\{\tilde{\beta}_3^{-1} n^{1/2}, \tilde{\gamma}_2^{-1} n\}, \quad a^{-1} = \max\{\tilde{\beta}_3 n^{-1/2}, \tilde{\gamma}_2 n^{-1}\}.$$

We may assume that $a > 404$. In case $a \le 404$ it is clear that

$$(\tfrac{96}{95})^2 \int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \le 404\, (\tfrac{96}{95})^2 \le 413,$$

so that the statement of the Theorem clearly follows from (2.4). Note that via Lyapunov's inequality (see Petrov (1995), p. 7)

$$\tilde{\beta}_3 \ge (s^2)^{3/2} = 1 \quad \text{and} \quad a \le \tilde{\beta}_3^{-1} n^{1/2} \le n^{1/2}. \tag{2.5}$$

To complete the proof, by (2.4) it will suffice to show that

$$\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \le 404. \tag{2.6}$$

From now on, let $c := 404$. The method to obtain (2.6) is as follows: we divide the interval $[0, a]$ into two parts, $[0, a^{1/4}]$ and $[a^{1/4}, a]$, and use different methods in order to show that the corresponding two integrals both are bounded by a constant. In order to do this we must subdivide the interval $[a^{1/4}, a]$ again into a few intervals, on all of which we may apply the same method. For $p < q$ such that $\frac{1}{4} < q \le 1$ and $p > q - \frac{1}{4}$, we set

$$\varepsilon_1 := \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt, \qquad \varepsilon_2 = \varepsilon_2(p, q) := \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt. \tag{2.7}$$

We estimate $\varepsilon_1$ and $\varepsilon_2$ from above in the Lemmas 2.3 and 2.4.

**Lemma 2.3.** *Writing* $c_1^2 := 1 - \frac{1}{3}c^{-3/4}$, *we have that*

$$\varepsilon_1 \;=\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\le\; 2^{-3/2} + (\tfrac{1}{2}\pi)^{1/2}\, c_1^{-1}.$$

**Proof of Lemma 2.3.** In order to estimate $\varepsilon_1$ we just remove the non-linear part $\mathbb{T}_2$ from $\mathbb{T} = \mathbb{T}_1 + \mathbb{T}_2$ by a Taylor expansion. As to the latter, we will use the following scheme. For any real or complex-valued function $f$ on $\mathbb{R}$, possessing a continuous $(k+1)^{\text{th}}$ derivative on some open interval containing both 0 and $x$, we have that

$$f(x) \;=\; \sum_{l=0}^{k} \tfrac{1}{l!}\, f^{(l)}(0)\, x^l + r_{k+1}(x) \tag{2.8}$$

where

$$r_{k+1}(x) \;:=\; \tfrac{1}{k!} \int_0^1 (1-s)^k\, f^{(k+1)}(sx)\, x^{k+1}\, ds.$$

Introducing a random variable $\tau$ which is uniformly distributed on $(0,1)$, we may also write

$$r_{k+1}(x) \;=\; \tfrac{1}{k!}\, \mathbb{E}\, (1-\tau)^k\, f^{(k+1)}(\tau x)\, x^{k+1}.$$

Using this we notice that

$$\begin{aligned}
\varepsilon_1 \;&=\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\}|\, dt \\
&\le\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}_1\}|\, dt + \int_0^{a^{1/4}} t\, \mathbb{E}\, |\mathbb{T}_2|\, dt \;=:\; \varepsilon_{11} + \varepsilon_{12},
\end{aligned}$$

where, with (A.17) and (A.24), $\mathbb{E}\, \mathbb{T}_2^2 \le \frac{1}{2}\tilde{\gamma}_2\, n^{-1} \le \frac{1}{2}a^{-1}$, so that

$$\varepsilon_{12} \;\le\; \tfrac{1}{2}a^{1/2}\, \{\mathbb{E}\, \mathbb{T}_2^2\}^{1/2} \;\le\; 2^{-3/2}.$$

With regard to $\varepsilon_{11}$ we need the following inequality:

$$|\mathbb{E}\, e\{tT_1\}| \;\le\; 1 - \tfrac{1}{2}t^2 n^{-1}\, c_1^2 \;\le\; \exp\{-\tfrac{1}{2}t^2 n^{-1}\, c_1^2\}. \tag{2.9}$$

Indeed, because $e\{tT_1\} = 1 + (it)\, T_1 + \frac{1}{2}(it)^2\, T_1^2 + R_1$ with $|R_1| \le \frac{1}{6}t^3\, |T_1|^3$ (again using (2.8)), we have that

$$\begin{aligned}
|\mathbb{E}\, e\{tT_1\} - (1 - \tfrac{1}{2}t^2 n^{-1})| \;&\le\; \tfrac{1}{6}t^3\, \mathbb{E}\, |T_1|^3 \\
&\le\; \tfrac{1}{6}t^2\, a^{1/4}\, n^{-1}\, (\tilde{\beta}_3\, n^{-1/2}) \;\le\; \tfrac{1}{6}t^2\, a^{-3/4}\, n^{-1},
\end{aligned}$$

while $1 - \frac{1}{2}t^2 n^{-1} \geq 0$ for $0 \leq t \leq a^{1/4}$ and $a \geq c$, and (2.9) is clearly correct. As a result of it we see that

$$\varepsilon_{11} = \int_0^{a^{1/4}} |\mathbb{E}\, e\{tT_1\}|^n \, dt \leq \int_0^\infty \exp\{-\tfrac{1}{2}c_1^2\, t^2\}\, dt = (\tfrac{1}{2}\pi)^{1/2}\, c_1^{-1},$$

which completes the proof. □

We turn to the estimation of $\varepsilon_2$.

**Lemma 2.4.** *Let $\frac{1}{4} < q \leq 1$ and $q - \frac{1}{4} < p < q$. Write*

$$r \;:=\; \frac{q - \frac{1}{4}}{p - (q - \frac{1}{4})} \quad \textit{and} \quad \tilde{c}^2 \;:=\; (1 - c^{2q-5/2})\,(1 - \tfrac{1}{3}c^{-(1-q)} - 2\,\delta^{1/2} - \delta)$$

*for any $\delta > 0$ such that this expression is positive. Then we have that*

$$\varepsilon_2(p,q) \;\leq\; 2^{-3/2} + \delta^{-1}\, c^{-(1-q)} + \tilde{c}^{-(1+r)} \int_0^\infty y^r \, \exp\{-\tfrac{1}{2}y^2\}\, dy.$$

**Proof of Lemma 2.4.** The proof of Lemma 2.4 is less straightforward. We start by splitting up the sample in two parts, $X_1, \ldots, X_m$ and $X_{m+1}, \ldots, X_n$, for some convenient $m$. In this way we split up $\mathbb{T} = \mathbb{T}_1 + \mathbb{T}_2$ by writing

$$\mathbb{T}_1 \;=\; \mathbb{T}_{11} + \mathbb{T}_{12} \;:=\; \textstyle\sum_{j=1}^m T_j + \sum_{j=m+1}^n T_j \tag{2.10}$$

and $\mathbb{T}_2 = \mathbb{T}_{21} + \mathbb{T}_{22} + \mathbb{T}_{23}$, with

$$\mathbb{T}_{21} := \textstyle\sum_{1 \leq j < k \leq m} T_{j,k}, \quad \mathbb{T}_{22} := \sum_{1 \leq j \leq m < k \leq n} T_{j,k}, \quad \mathbb{T}_{23} := \sum_{m < j < k \leq n} T_{j,k}. \tag{2.11}$$

We choose the number $m$ of the form $[n\, a^{-r}]$, for some $r \in [0, 2]$, in such a way that it is as large as possible, while still allowing us to remove $\mathbb{T}_{21}$ by a Taylor expansion. After this removal has taken place we condition the statistic on $X_{m+1}, \ldots, X_n$, which we will from here on denote by

$$Y \;:=\; (X_{m+1}, \ldots, X_n). \tag{2.12}$$

The parts of $\mathbb{T}$ that still matter at this moment are $\mathbb{T}_{11}$ and $\mathbb{T}_{22}$, which conditionally are sums of independent random variables, $\mathbb{T}_{11}$ being $T_1 + \ldots + T_m$ with $T_j = T_j(X_j)$, and $\mathbb{T}_{22} = M_1 + \ldots + M_m$ with

$$M_j \;=\; M_j(X_j, Y) \;:=\; T_{j,m+1} + \ldots + T_{j,n}, \quad \text{for } 1 \leq j \leq m. \tag{2.13}$$

To smoothen the proof we need a truncation argument. Let

$$\varphi \;=\; \varphi(Y) \;:=\; \mathbb{E}\,(M_1^2 \,|\, Y). \tag{2.14}$$

Take any (small) $\delta > 0$ and let $I\{\varphi\} := I\{\varphi \le \delta\, n^{-1}\}$. We condition on the event that $I\{\varphi\} = 1$. To this, note that, by Markov's inequality,

$$\begin{aligned}\mathbb{P}(\varphi > \delta\, n^{-1}) \;&\le\; \delta^{-1} n\, \mathbb{E}\,\varphi \;=\; \delta^{-1}\, n\,(n-m)\,\mathbb{E}\,T_{1,2}^2 \\ &\le\; \delta^{-1}\,(\tilde\gamma_2\, n^{-1}) \;\le\; \delta^{-1}\, a^{-1}.\end{aligned} \tag{2.15}$$

As a result, we see that

$$\begin{aligned}\int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;&\le\; \int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\}|\, dt + \int_{a^p}^{a^q} |\mathbb{E}\,(1 - I\{\varphi\})\, e\{t\mathbb{T}\}|\, dt \\ &\le\; \int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\}|\, dt + \delta^{-1}\, a^{-1+q},\end{aligned} \tag{2.16}$$

as

$$|\mathbb{E}\,(1 - I\{\varphi\})\, e\{t\mathbb{T}\}| \;\le\; \mathbb{E}\,(1 - I\{\varphi\}) \;=\; \mathbb{P}(\varphi > \delta\, n^{-1}) \;\le\; \delta^{-1}\, a^{-1+q}.$$

Now let $\tilde M_j := I\{\varphi\}\, M_j$ for $1 \le j \le m$, $\tilde{\mathbb{T}}_{22} := I\{\varphi\}\, \mathbb{T}_{22} = \sum_{j=1}^m \tilde M_j$ and

$$\tilde{\mathbb{T}} \;:=\; \mathbb{T}_{11} + \mathbb{T}_{12} + \mathbb{T}_{21} + \tilde{\mathbb{T}}_{22} + \mathbb{T}_{23}\,.$$

By construction $\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\} = \mathbb{E}\, I\{\varphi\}\, e\{t\tilde{\mathbb{T}}\}$, so that from here on we may concentrate on

$$\varepsilon_3 \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t\tilde{\mathbb{T}}\}|\, dt.$$

Now take

$$m \;:=\; [n\, a^{-2q+1/2}],$$

noting that $1 \le m \le n$. We remove $\mathbb{T}_{21}$ by means of a Taylor expansion. As $\mathbb{E}\,\mathbb{T}_{21}^2 = (mn^{-1})^2\, \frac{1}{2}\,(\tilde\gamma_2\, n^{-1}) \le \frac{1}{2} a^{-4q+1}\, a^{-1} = \frac{1}{2} a^{-4q}$,

$$\varepsilon_3 \;\le\; \int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t(\tilde{\mathbb{T}} - \mathbb{T}_{21})\}|\, dt + \int_{a^p}^{a^q} t\, \mathbb{E}\,|\mathbb{T}_{21}|\, dt \;=:\; \varepsilon_{31} + \varepsilon_{32},$$

with $\varepsilon_{32} \le \frac{1}{2} a^{2q}\, \{\mathbb{E}\,\mathbb{T}_{21}^2\}^{1/2} \le 2^{-3/2}$. As to $\varepsilon_{31}$, using a Taylor expansion we see that

$$|\mathbb{E}\,(e\{t(T_1 + \tilde M_1)\}\,|\,Y)| \;\le\; |\mathbb{E}\,(e\{tT_1\} + (it)\, e\{tT_1\}\tilde M_1 \,|\, Y)| + \tfrac{1}{2} t^2\, \mathbb{E}\,(\tilde M_1^2 \,|\, Y).$$

Now by construction $\mathbb{E}(\tilde{M}_1^2 \,|\, Y) = I\{\varphi\}\,\varphi \le \delta\, n^{-1}$. Furthermore,

$$e\{tT_1\} \;=\; 1 + R_1 \tag{2.17}$$

for a certain $R_1 = R_1(X_1)$ with $|R_1| \le t\,|T_1|$, and moreover

$$\mathbb{E}(\tilde{M}_1 \,|\, Y) \;=\; I\{\varphi\}\,\mathbb{E}(M_1 \,|\, Y) \;=\; 0,$$

so that with Hölder's inequality

$$\begin{aligned} |\mathbb{E}(e\{tT_1\}\tilde{M}_1 \,|\, Y)| \;&=\; |\mathbb{E}(R_1\,\tilde{M}_1 \,|\, Y)| \\ &\le\; t\,\mathbb{E}(|T_1\tilde{M}_1| \,|\, Y) \;\le\; n^{-1/2}\,(\delta\, n^{-1})^{1/2}\, t \;=\; \delta^{1/2}\, t\, n^{-1}. \end{aligned}$$

As in (2.9), we have that $|\mathbb{E}\, e\{tT_1\}| \le 1 - \frac{1}{2}(1 - \frac{1}{3}c^{-(1-q)})\, t^2\, n^{-1}$, and as a result, for any $t \in [a^p, a^q]$,

$$\begin{aligned} |\mathbb{E}(e\{t(T_1 + \tilde{M}_1)\} \,|\, Y)| \;&\le\; 1 - \tfrac{1}{2}(1 - \tfrac{1}{3}c^{-(1-q)} - 2\,\delta^{1/2} - \delta)\, t^2\, n^{-1} \\ &\le\; \exp\{-\tfrac{1}{2}\hat{c}\, t^2\, n^{-1}\}, \end{aligned} \tag{2.18}$$

where we write $\hat{c} := 1 - \frac{1}{3}c^{-(1-q)} - 2\,\delta^{1/2} - \delta$. It is now easily seen that

$$\begin{aligned} \varepsilon_{31} \;&\le\; \textstyle\int_{a^p}^{a^q} \mathbb{E}\,|I\{\varphi\}\,\mathbb{E}(e\{t(\tilde{\mathbb{T}} - \mathbb{T}_{21})\} \,|\, Y)|\, dt \\ &=\; \textstyle\int_{a^p}^{a^q} \mathbb{E}\, I\{\varphi\}\, |e\{t(\mathbb{T}_{12} + \mathbb{T}_{23})\}|\, |\mathbb{E}(e\{t(\mathbb{T}_{11} + \tilde{\mathbb{T}}_{22})\} \,|\, Y)|\, dt \\ &\le\; \textstyle\int_{a^p}^{a^q} \mathbb{E}\,|\mathbb{E}(e\{t(T_1 + \tilde{M}_1)\} \,|\, Y)|^m\, dt \\ &\le\; \textstyle\int_{a^p}^{a^q} \exp\{-\frac{1}{2}\hat{c}\, t^2\, mn^{-1}\}\, dt. \end{aligned} \tag{2.19}$$

Since

$$\begin{aligned} mn^{-1} \;&\ge\; a^{-2q+1/2} - n^{-1} \\ &\ge\; a^{-2q+1/2}\,(1 - a^{2q-5/2}) \;\ge\; a^{-2q+1/2}\,(1 - c^{2q-5/2}), \end{aligned}$$

(note that by (2.5) we have that $n^{-1} \le a^{-2}$), it follows that

$$\begin{aligned} \varepsilon_{31} \;&\le\; \textstyle\int_{a^p}^{a^q} \exp\{-\frac{1}{2}\tilde{c}^2\, a^{-2q+1/2}\, t^2\}\, dt \\ &\le\; (\tilde{c}^2\, a^{-2q+1/2})^{-1/2} \textstyle\int_{(\tilde{c}^2\, a^{-2q+1/2})^{1/2}\, a^p}^{\infty} \exp\{-\frac{1}{2}y^2\}\, dy \\ &=\; \tilde{c}^{-1}\, a^{q-1/4} \textstyle\int_{\tilde{c}\, a^{p-q+1/4}}^{\infty} y^{-r}\, y^r\, \exp\{-\frac{1}{2}y^2\}\, dy \\ &\le\; \tilde{c}^{-(1+r)} \textstyle\int_0^{\infty} y^r\, \exp\{-\frac{1}{2}y^2\}\, dy \end{aligned} \tag{2.20}$$

(note that the final inequality is the point where we need the assumption that $p > q - \frac{1}{4}$), which finishes the proof. $\square$

The proof of Theorem 2.1 for $U$-statistics is now easily concluded by splitting the interval $[a^{1/4}, a]$ into several intervals of type $[a^p, a^q]$, and applying Lemma 2.4. As the result of Theorem 2.1 is stated for arbitrary statistics, we postpone calculations to the next section.

## 2.3 Proof of Theorem 2.1 for arbitrary statistics

We go about in the same way as for $U$-statistics of order 2. Again we may assume without loss of generality that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We put

$$a := \min\{\tilde{\beta}_3^{-1}\, n^{1/2},\, \tilde{\Delta}_2^{-2}\, n\}, \quad a^{-1} = \max\{\tilde{\beta}_3\, n^{-1/2},\, \tilde{\Delta}_2^2\, n^{-1}\},$$

and prove the theorem via inequality (2.4). Again we assume that $a > c$ and again we will show that

$$\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\le\; c.$$

In order to do this we need to decompose $\mathbb{T}$ in the following way. Let $m \in N$. As earlier we write

$$\mathbb{T}_{11} \;:=\; \sum_{j=1}^m T_j.$$

For $1 \le j \le m$ we write

$$M_j := \sum_{k=m+1}^n T_{j,k} \quad \text{and} \quad N_j := \sum_{A:\,|A|\ge 3,\, A\cap\{1,\dots,m\}=\{j\}} T_A. \tag{2.21}$$

Furthermore, let

$$\Lambda \;=\; \Lambda(m) \;:=\; \sum_{A:\,|A\cap\{1,\dots,m\}|\ge 2} T_A, \tag{2.22}$$

and $R := \mathbb{E}\,(\mathbb{T}\,|\,X_{m+1}, \dots, X_n) = \sum_{A\subset\{m+1,\dots,n\}} T_A$. Now

$$\mathbb{T} \;=\; \mathbb{T}_{11} + \sum_{j=1}^m (M_j + N_j) + \Lambda + R, \tag{2.23}$$

and we have divided up $\mathbb{T}$'s Hoeffding decomposition into groups of $T_A$ according to the fact whether the $A$ have either zero, one, or more than one elements in common with $\{1, \dots, m\}$. Divided up in this way we obtain

$$R, \quad \mathbb{T}_{11} + \textstyle\sum_{j=1}^{n}(M_j + N_j) \quad \text{and} \quad \Lambda$$

respectively. In the case of $U$-statistics of order 2, for example, we have:

$$\Lambda = 0, \quad \textstyle\sum_{j=1}^{m} M_j = \mathbb{T}_{22}, \quad \textstyle\sum_{j=1}^{m} N_j = 0 \quad \text{and} \quad R = \mathbb{T}_{12} + \mathbb{T}_{23}.$$

Lemma 2.5 will provide us with bounds for the second moments $\mathbb{E}\,\Lambda^2$ and $\mathbb{E}\,N_1^2$. Our strategy is now as follows: as concerns $\varepsilon_1$ we just take $m = n$ after which we easily remove $\Lambda = \sum_{|A|\geq 3} T_A$. As concerns $\varepsilon_2(p,q)$ we take $m$ as before and remove $\Lambda$ by a Taylor expansion. After that we take $M_j + N_j$ instead of $M_j$ in the original proof and use the same truncation.

The lemma reads as follows (cf. Bentkus, Götze and Van Zwet (1997), Lemma 4.1):

**Lemma 2.5.** *We have that*

$$\mathbb{E}\,\Lambda^2 \;\leq\; \tfrac{1}{2}(mn^{-1})^2\,(\tilde{\Delta}_2^2\,n^{-1}) \quad \textit{and} \quad \mathbb{E}\,N_1^2 \;\leq\; n^{-1}\,(\tilde{\Delta}_2^2\,n^{-1}).$$

**Proof of Lemma 2.5.** We may assume that $m \geq 2$. Using (A.23), we see that

$$\begin{aligned} \mathbb{E}\,\Lambda^2 &= \textstyle\sum_{A:\,|A\cap\{1,\dots,m\}|\geq 2} \mathbb{E}\,T_A^2 \;\leq\; \sum_{1\leq j<k\leq m} \sum_{j,k\in A} \mathbb{E}\,T_A^2 \\ &= \textstyle\binom{m}{2} \sum_{\{1,2\}\subset A} \mathbb{E}\,T_A^2 \;\leq\; \tfrac{1}{2} m^2\, n^{-2}\,(\tilde{\Delta}_2^2\,n^{-1}). \end{aligned}$$

As to $\mathbb{E}\,N_1^2$, we may suppose that $m \notin \{n-1, n\}$, since otherwise $N_1 = 0$. Using again (A.23) we see that

$$\begin{aligned} \mathbb{E}\,N_1^2 &= \textstyle\sum_{A:\,|A|\geq 2,\, A\subset\{m+1,\dots,n\}} \mathbb{E}\,T_{1,A}^2 \\ &\leq (n-m) \textstyle\sum_{\emptyset\neq A\subset\{m+2,\dots,n\}} \mathbb{E}\,T_{1,m+1,A}^2 \\ &\leq n \textstyle\sum_{A\subset\{3,\dots,n\}} \mathbb{E}\,T_{1,2,A}^2 \;=\; n^{-1}\,(\tilde{\Delta}_2^2\,n^{-1}). \end{aligned} \tag{2.24}$$

This finishes the proof. $\square$

Now we proceed as promised. Let $\varepsilon_1$ and $\varepsilon_2(p,q)$ be as in (2.7). First we look at $\varepsilon_1$. Let $\mathbb{T}_1 := \sum_{j=1}^{n} T_j$ and $\mathbb{T}_2 := \sum_{A:\,|A|=2} T_A$ as before. Taking

$m = n$ we have that $\Lambda = \sum_{A:\,|A|\geq 3} T_A$, while by Lemma 2.5 it is clear that $\mathbb{E}\,\Lambda^2 \leq \frac{1}{2}\tilde{\Delta}_2^2\, n^{-1} \leq \frac{1}{2}a^{-1}$. Therefore

$$\begin{aligned}\int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\} - \mathbb{E}\,e\{t(\mathbb{T}-\Lambda)\}|\,dt \;&\leq\; \int_0^{a^{1/4}} t\,\mathbb{E}\,|\Lambda|\,dt \\ &\leq\; (\tfrac{1}{2}a^{-1})^{1/2}\,\tfrac{1}{2}a^{1/2} \;=\; 2^{-3/2},\end{aligned}$$

and, via Lemma 2.3,

$$\varepsilon_1 \;\leq\; 2^{-3/2} + 2^{-3/2} + (\tfrac{1}{2}\pi)^{1/2}\,(1-\tfrac{1}{3}c^{-3/4})^{-1/2}.$$

As to $\varepsilon_2(p,q)$, we take $m := [n\,a^{-2q+1/2}]$ as earlier. The cost of removing $\Lambda$ from $\mathbb{T}$ is (see Lemma 2.5) bounded by

$$\begin{aligned}\int_{a^p}^{a^q} |\mathbb{E}\,e\{t\mathbb{T}\} - \mathbb{E}\,e\{t(\mathbb{T}-\Lambda)\}|\,dt \;&\leq\; \int_{a^p}^{a^q} t\,\{\mathbb{E}\,\Lambda^2\}^{1/2}\,dt \\ &\leq\; \tfrac{1}{2}a^{2q}\,\{\tfrac{1}{2}(mn^{-1})^2\,(\tilde{\Delta}_2^2\,n^{-1})\}^{1/2} \\ &\leq\; 2^{-3/2}\,a^{2q}\,a^{-2q+1/2}\,a^{-1/2} \;=\; 2^{-3/2}.\end{aligned} \tag{2.25}$$

Now again take $Y := (X_{m+1},\dots,X_n)$ and set

$$\varphi \;=\; \varphi(Y) \;:=\; \mathbb{E}\big((M_1+N_1)^2\,|\,Y\big). \tag{2.26}$$

By Lemma 2.5 and the fact that, with (A.23), $\mathbb{E}\,M_1^2 = (n-m)\,\mathbb{E}\,T_{1,2}^2 \leq n^{-1}\,(\tilde{\Delta}_2^2\,n^{-1})$,

$$\mathbb{E}\,\varphi \;\leq\; 2n^{-1}\,(\tilde{\Delta}_2^2\,n^{-1}) \;\leq\; 2n^{-1}\,a^{-1},$$

so replacement of

$$\hat{\mathbb{T}} \;:=\; \mathbb{T} - \Lambda \;=\; \mathbb{T}_{11} + \textstyle\sum_{j=1}^m (M_j+N_j) + R$$

by the truncated version

$$\tilde{\mathbb{T}} \;:=\; \mathbb{T}_{11} + I\{\varphi \leq \delta\,n^{-1}\}\,\textstyle\sum_{j=1}^m (M_j+N_j) + R$$

will cost us (see (2.16)) $\leq 2\delta^{-1}\,a^{-(1-q)} \leq 2\delta^{-1}\,c^{-(1-q)}$. After this we estimate $\varepsilon_2$ in completely the same way as $\varepsilon_{31}$ in the proof of Lemma 2.4. This leads us to:

$$\varepsilon_2(p,q) \;\leq\; 2^{-3/2} + 2^{-3/2} + 2\delta^{-1}\,c^{-(1-q)} + \tilde{c}^{\,-(1+r)}\int_0^\infty y^r\,\exp\{-\tfrac{1}{2}y^2\}\,dy,$$

with $\tilde{c}$ and $r$ as in Lemma 2.4.

Finally we turn to the concrete calculations. What we need is a sequence of intervals $[p_j, q_j]$, for $1 \le j \le k$, such that $q_1 = 1$, $q_{j+1} = p_j$ $(1 \le j \le k-1)$ and $p_k \le \frac{1}{4}$, after which

$$\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\le\; \varepsilon_1 + \sum_{j=1}^k \varepsilon_2(p_j, q_j)$$

and we are able to prove (2.6). We make the situation easier to handle by always taking $r = 4$ and $\delta = 0.2$. This results in the sequence

$$p_1 = \tfrac{15}{16}, \quad p_2 = \tfrac{55}{64}, \quad p_3 = \tfrac{195}{264}, \quad p_4 = \tfrac{655}{1024}, \quad p_5 = \tfrac{1995}{4096}, \quad p_6 = \tfrac{4855}{16384}, \quad p_7 \le \tfrac{1}{4}.$$

Notice that

$$\int_0^\infty y^4 \exp\{-\tfrac{1}{2}y^2\}\, dy \;=\; 3\,(\tfrac{1}{2}\pi)^{1/2},$$

so that we get to

$$\begin{aligned}\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\;\le\;\; & 16 \cdot 2^{-3/2} + (\tfrac{1}{2}\pi)^{1/2}\,(1 - \tfrac{1}{3}c^{-3/4})^{-1/2} \\ & + 3\,(\tfrac{1}{2}\pi)^{1/2} \sum_{j=1}^7 \tilde{c}_j^{-5} + 2\delta^{-1} \sum_{j=1}^7 c^{-(1-q_j)},\end{aligned}$$

which in turn leads us to (2.6). This finishes the proof of Theorem 2.1 in the case that $s$ is the norming factor.

We conclude by giving the proof of Theorem 2.1 in the case that $\sigma$ is the norming factor. Here we have that $\sigma^2 = s^2 + R^2$, where (see (A.17) and (A.24))

$$R^2 \;\le\; \tfrac{1}{2}\,\tilde{\Delta}_2^2\, n^{-1}. \tag{2.27}$$

In the case that $s^2/\sigma^2 \ge \frac{709}{710}$ the result of the theorem is now easily derived from the equality

$$Q(\mathbb{T}/\sigma, \lambda) \;=\; Q(\mathbb{T}/s, (\sigma/s)\,\lambda),$$

together with the result already obtained. On the other hand: in the case that $s^2/\sigma^2 < \frac{709}{710}$ it is clear that $R^2/\sigma^2 \ge \frac{1}{710}$, so that, with (2.27), $\frac{1}{2}\tilde{\Delta}_2^2\,\sigma^{-2}\, n^{-1} \ge \frac{1}{710}$. But then the desired result follows from the trivial estimate

$$Q(\mathbb{T}/\sigma, \lambda) \;\le\; 1.$$

This concludes the proof of Theorem 2.1. $\square$

## 2.4 Proof of Theorem 2.2 for $U$-statistics of order 2

We turn to the proof of Theorem 2.2. First we look at the case of $U$-statistics of order 2. Here we have that $\Delta^2 = \Delta_1^2 = \operatorname{var}\mathbb{T}_2$, cf. (A.16). As in Section 2.2, we assume that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We set

$$a \ := \ \min\{\beta^{-1}, \Delta_1^{-2}\}, \quad a^{-1} \ = \ \max\{\beta, \Delta_1^2\}$$

and prove that $\int_0^a |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt$ is bounded by some absolute constant. We may assume that $a \geq 100$.

We pay no attention to concrete values of constants anymore: the reason for this is that the constants involved are getting too big to be of practical use anymore. We use the symbol $\ll$ as in (2.1), and are going to prove that $\int_0^a |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ll 1$.

We proceed as earlier. First we will show that as in Section 2.2

$$\varepsilon_1 \ := \ \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ll 1.$$

After that we prove that

$$\varepsilon_2 \ = \ \varepsilon_2(p,q) \ := \ \int_{a^p}^{a^q} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ \ll_{p,q} \ 1,$$

for any $\frac{1}{4} < q \leq 1$ and $q - \frac{1}{4} < p \leq q$, and the result of the theorem is an easy consequence. We start with $\varepsilon_1$:

**Lemma 2.6.** *We have that*

$$\varepsilon_1 \ = \ \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ll 1.$$

**Proof of Lemma 2.6.** A simple Taylor expansion is telling us that $\varepsilon_1 \leq \varepsilon_{11} + \varepsilon_{12}$ with

$$\varepsilon_{11} \ := \ \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}_1\}|\,dt$$

and

$$\varepsilon_{12} \ := \ \int_0^{a^{1/4}} t\,\mathbb{E}\,|\mathbb{T}_2|\,dt \ \leq \ \tfrac{1}{2}\Delta_1\,a^{1/2} \leq \tfrac{1}{2},$$

so we only need to find a bound for $\varepsilon_{11}$. As to $\varepsilon_{11}$ we have the following inequality:

$$|\mathbb{E}\, e\{tT_j\}| \ \le\ \exp\{-\tfrac{1}{2}t^2\,(s_j^2 - \tfrac{1}{3}\beta_j\,\beta^{-1/4})\}. \tag{2.28}$$

Indeed, since $s_j^2 \le \beta_j^{2/3} \le \beta^{2/3}$ we have that $1 - \frac{1}{2}s_j^2\,t^2 \ge 0$ for $t \in [0, a^{1/4}]$. The Taylor expansion

$$e\{tT_j\} \ =\ 1 + (it)\,T_j + \tfrac{1}{2}(it)^2\,T_j^2 + R_j,$$

with $|R_j| \le \frac{1}{6}\,t^3\,|T_j|^3$, now shows us that

$$\mathbb{E}\, e\{tT_j\} \ =\ 1 - \tfrac{1}{2}s_j^2\,t^2 + \mathbb{E}\, R_j\,,$$

so that $|\mathbb{E}\, e\{tT_j\}| \le 1 - \frac{1}{2}s_j^2\,t^2 + \frac{1}{6}\,\beta_j\,t^3$. Using moreover that $1 - y \le \exp\{-y\}$ for all $y \in \mathbb{R}$, from this (2.28) easily follows. Now we note that as $a \ge 100$ we have that $\beta \le \frac{1}{100}$, and (2.28) leads us to:

$$\begin{aligned} \varepsilon_{11} \ &=\ \int_0^{a^{1/4}} \prod_{j=1}^n |\mathbb{E}\, e\{tT_j\}|\, dt \\ &\le\ \int_0^{a^{1/4}} \exp\{-\tfrac{1}{2}t^2\, \textstyle\sum_{j=1}^n (s_j^2 - \tfrac{1}{3}\beta_j\,\beta^{-1/4})\}\, dt \\ &=\ \int_0^{a^{1/4}} \exp\{-\tfrac{1}{2}(1 - \tfrac{1}{3}\beta^{3/4})\,t^2\}\, dt \ \ll\ 1, \end{aligned}$$

which concludes the proof of the Lemma. □

**Lemma 2.7.** *Let $q \in (\frac{1}{4}, 1]$ and $p \in (q - \frac{1}{4}, q)$. Then we have that*

$$\varepsilon_2 \ =\ \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \ \ \ll_{p,q}\ \ 1.$$

**Proof of Lemma 2.7.** In the i.i.d. case we have treated the integral $\varepsilon_2$ by splitting up the sample into two parts, $X_1, \ldots, X_m$ versus $X_{m+1}, \ldots, X_n$, where $m$ is of the form $[n\,a^{-r}]$ for some $r \in [0, 2]$. In this way $\mathbb{T}$ is split up as

$$\mathbb{T} \ =\ \mathbb{T}_1 + \mathbb{T}_2 \ =\ \mathbb{T}_{11} + \mathbb{T}_{12} + \mathbb{T}_{21} + \mathbb{T}_{22} + \mathbb{T}_{23}$$

as in (2.10) and (2.11). Now $\mathbb{T}_{21}$ is removed by a Taylor expansion and using conditional expectations with respect to $X_{m+1}, \ldots, X_n$ we get to the result that we want. In the present case this approach is not satisfactory as in a sense it privileges the first $m$ random variables over the remaining $n-m$. This problem is overcome by means of a very useful technique taken from Bentkus, Bloznelis and Götze (1996), to which we will refer as 'randomization'.

To this, let $\overline{X} := (\overline{X}_1, \dots, \overline{X}_n)$ be an independent copy of $X_1, \dots, X_n$, and let $\alpha_1, \dots, \alpha_n$ be an i.i.d. sample from a Bernoulli distributed population with

$$\mathbb{P}(\alpha_1 = 1) = 1 - \mathbb{P}(\alpha_1 = 0) = m,$$

for some $m \in [0,1]$ of the form $a^{-r}$ $(r \geq 0)$. The sample $\alpha := (\alpha_1, \dots, \alpha_n)$ is chosen independently of all other random variables involved. In particular we take

$$m \ := \ a^{-(2q-\frac{1}{2})}.$$

Now we replace the original sample $X_1, \dots, X_n$ by the 'mixed' sample

$$\alpha_1 X_1 + (1-\alpha_1)\overline{X}_1, \ \dots \ , \alpha_n X_n + (1-\alpha_n)\overline{X}_n.$$

Clearly the two samples are identically distributed, as this is so for each fixed outcome of $\alpha$. Thus, letting

$$\tilde{\mathbb{T}} \ := \ \mathbb{T}\big(\alpha_1 X_1 + (1-\alpha_1)\overline{X}_1, \ \dots \ , \ \alpha_n X_n + (1-\alpha_n)\overline{X}_n\big),$$

we have that $\mathbb{T} \stackrel{d}{=} \tilde{\mathbb{T}}$, where by $\stackrel{d}{=}$ we mean equality in distribution. As a consequence $\mathbb{E}\, e\{t\tilde{\mathbb{T}}\} = \mathbb{E}\, e\{t\mathbb{T}\}$, so it does not matter whether we replace $\mathbb{T}$ by $\tilde{\mathbb{T}}$.

Instead of isolating the first $m$ variables $X_1, \dots, X_m$ from the original sample, we may now take the set of $X_j$ for which $\alpha_j = 1$ from the mixed sample. Our gain is that in a sense a randomization of the choices of the $m$ variables has taken place.

We wonder what the Hoeffding decomposition of $\tilde{\mathbb{T}}$ looks like. To this, look at the statistic $\tilde{\mathbb{T}}$ for some fixed outcome of $\alpha$. As an example we take the case in which $\alpha = (0, 1, \dots, 1)$, that is, the case in which we look at the mixed sample $\overline{X}_1, X_2, \dots, X_n$. Here $\tilde{\mathbb{T}} = \mathbb{T}(\overline{X}_1, X_2, \dots, X_n)$, with a Hoeffding decomposition of the form

$$\tilde{T}_1(\overline{X}_1) + \textstyle\sum_{j=2}^n \tilde{T}_j(X_j) + \sum_{j=2}^n \tilde{T}_{1,j}(\overline{X}_1, X_j) + \sum_{2 \leq j < k \leq n} \tilde{T}_{j,k}(X_j, X_k),$$

with the $\tilde{T}_j$ and the $\tilde{T}_{j,k}$ as in (A.4), using the mixed sample. Since $T_j$ and $\tilde{T}_j$ are functions depending only on the distribution of the underlying sample, and the original and the mixed sample are identically distributed, we may take $T_j = \tilde{T}_j$ for all $j$. In the same way we may take $T_A = \tilde{T}_A$ for any $A \subset N$. As a result, the random vector

$$\big(\tilde{T}_1(\overline{X}_1),\, \tilde{T}_2(X_2),\, \dots\,,\, \tilde{T}_n(X_n),\, \tilde{T}_{1,2}(\overline{X}_1, X_2),\, \dots\,,\, \tilde{T}_{n-1,n}(X_{n-1}, X_n)\big) \tag{2.29}$$

is equal to

$$\big(T_1(\overline{X}_1),\, T_2(X_2),\, \dots\,,\, T_n(X_n),\, T_{1,2}(\overline{X}_1, X_2),\, \dots\,,\, T_{n-1,n}(X_{n-1}, X_n)\big),$$

and therefore has the same distribution as $(T_1(X_1), \dots, T_{n-1,n}(X_{n-1}, X_n))$. Using the same type of argument, we see that in general

$$\tilde{\mathbb{T}} \;:=\; \tilde{\mathbb{T}}_1 + \tilde{\mathbb{T}}_2, \tag{2.30}$$

with

$$\tilde{\mathbb{T}}_1 \;=\; \tilde{\mathbb{T}}_{11} + \tilde{\mathbb{T}}_{12} \;:=\; \textstyle\sum_{j=1}^n \alpha_j\, T_j(X_j) + \sum_{j=1}^n (1-\alpha_j)\, T_j(\overline{X}_j), \tag{2.31}$$

and $\tilde{\mathbb{T}}_2 = \tilde{\mathbb{T}}_{21} + \tilde{\mathbb{T}}_{22} + \tilde{\mathbb{T}}_{23}$ with

$$\begin{aligned} \tilde{\mathbb{T}}_{21} &:= \textstyle\sum_{1\le j<k\le n} \alpha_j \alpha_k\, T_{j,k}(X_j, X_k), \\ \tilde{\mathbb{T}}_{22} &:= \textstyle\sum_{1\le j<k\le n} \alpha_j (1-\alpha_k)\, T_{j,k}(X_j, \overline{X}_k) \quad \text{and} \\ \tilde{\mathbb{T}}_{23} &:= \textstyle\sum_{1\le j<k\le n} (1-\alpha_j)(1-\alpha_k)\, T_{j,k}(\overline{X}_j, \overline{X}_k), \end{aligned} \tag{2.32}$$

where, e.g., by $T_{4,3}(X_4, \overline{X}_3)$ we actually mean $T_{3,4}(\overline{X}_3, X_4)$.

We continue with our estimation of $\varepsilon_2 = \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\tilde{\mathbb{T}}\}|\, dt$. For reasons of convenience we drop the tilde from our notation, so that we will just talk about $\mathbb{T}$, $\mathbb{T}_1$ and so on for the decomposition as defined in (2.30), (2.31) and (2.32).

First we get rid of $\mathbb{T}_{21}$: by a Taylor expansion we see that $\varepsilon_2 \le \varepsilon_{21} + \varepsilon_{22}$ with

$$\varepsilon_{21} \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t(\mathbb{T} - \mathbb{T}_{21})\}|\, dt \quad \text{and} \quad \varepsilon_{22} \;:=\; \int_{a^p}^{a^q} t\, \mathbb{E}\, |\mathbb{T}_{21}|\, dt,$$

where it is clear that

$$\varepsilon_{22} \;\le\; \tfrac{1}{2}\, \{m^2\, \Delta_1^2\}^{1/2}\, a^{2q} \;\le\; \tfrac{1}{2} a^{-(2q-\frac{1}{2})}\, a^{-1/2}\, a^{2q} \;=\; \tfrac{1}{2},$$

so that we only need to prove that $\varepsilon_{21} \ll_{p,q} 1$. To this, let

$$M_j := \sum_{k:k\neq j}(1-\alpha_k)\, T_{j,k}(X_j, \overline{X}_k) \quad \text{and} \quad \varphi_j^2 := \mathbb{E}\,(M_j^2 \,|\, \overline{X}, \alpha) \tag{2.33}$$

be the expressions corresponding to (2.13) and (2.14). Note that as in (2.19)

$$\varepsilon_{21} \;\le\; \int_{a^p}^{a^q} \mathbb{E} \prod_{j=1}^n V_j\, dt \tag{2.34}$$

with $V_j := |\mathbb{E}\,(e\{t\alpha_j (T_j + M_j)\} \,|\, \overline{X}, \alpha)|$. As in Section 2.2 we proceed by Taylor expanding, but we need to make the situation easier to handle. To this, in the case that $s_j^2 < 2^{-1/2}\,\beta_j\,\beta^{-1}$ we use the trivial bound $V_j \le 1$. Moreover, as before we condition on the event that $\sum_{j=1}^n \alpha_j\,\varphi_j^2 \le \delta\, m$ for some, small, $\delta > 0$. To this end, we set

$$\gamma_j \;:=\; I\{s_j^2 \ge 2^{-1/2}\,\beta_j\,\beta^{-1}\}, \qquad \tilde\alpha_j \;:=\; \gamma_j\,\alpha_j$$

and

$$I\{\varphi\} \;:=\; I\{\textstyle\sum_{j=1}^n \alpha_j\,\varphi_j^2 \le \delta\, m\}. \tag{2.35}$$

Moreover we set

$$W_j \;:=\; V_j^{\tilde\alpha_j} \;=\; |\mathbb{E}\,(e\{t\tilde\alpha_j (T_j + M_j)\} \,|\, \overline{X}, \alpha)|,$$

for which $V_j \le W_j \le 1$, and

$$\varepsilon_{23} \;:=\; \int_{a^p}^{a^q} \mathbb{E}\, I\{\varphi\} \prod_{j=1}^n W_j\, dt.$$

It is clear that

$$\varepsilon_{21} \;\le\; \int_{a^p}^{a^q} \mathbb{E} \prod_{j=1}^n W_j\, dt \;\le\; \varepsilon_{23} + \int_{a^p}^{a^q} \mathbb{E}\,(1 - I\{\varphi\})\, dt. \tag{2.36}$$

Markov's inequality leads us to the fact that

$$\begin{aligned} \mathbb{E}\,(1 - I\{\varphi\}) \;&=\; \mathbb{P}(\textstyle\sum_{j=1}^n \alpha_j\,\varphi_j^2 > \delta\, m) \;\le\; \delta^{-1}\, m^{-1}\, \mathbb{E} \textstyle\sum_{j=1}^n \alpha_j\,\varphi_j^2 \\ &\le\; \delta^{-1}\, m^{-1}\, m \textstyle\sum_{j=1}^n \mathbb{E}\, M_j^2 \;\le\; 2\delta^{-1}\,\Delta_1^2, \end{aligned} \tag{2.37}$$

so that $\int_{a^p}^{a^q} \mathbb{E}\,(1 - I\{\varphi\})\, dt \le 2\delta^{-1}\, a^{-(1-q)}$ and

$$\varepsilon_{21} \;\le\; \varepsilon_{23} + 2\delta^{-1}.$$

We prove that $\varepsilon_{23} \ll_{p,q} 1$.

As to $W_j$, expansions into Taylor series lead us to

$$\begin{aligned}
W_j &\le |\mathbb{E}\,(e\{t\tilde\alpha_j T_j\} + (it)\,e\{t\tilde\alpha_j T_j\}\,\tilde\alpha_j M_j \,|\, \overline{X}, \alpha)| + \tfrac{1}{2}t^2\,\tilde\alpha_j\,\mathbb{E}\,(M_j^2 \,|\, \overline{X}, \alpha) \\
&\le |\mathbb{E}\,(e\{t\tilde\alpha_j T_j\} \,|\, \alpha_j)| + t^2\,\tilde\alpha_j\,\mathbb{E}\,(|T_j\,M_j| \,|\, \overline{X}, \alpha) + \tfrac{1}{2}t^2\,\tilde\alpha_j\,\varphi_j^2 \\
&\le |1 - \tfrac{1}{2}t^2\,\tilde\alpha_j\,s_j^2| + \tfrac{1}{6}t^3\,\tilde\alpha_j\,\beta_j \\
&\qquad + t^2\,\tilde\alpha_j\,\mathbb{E}\,(|\delta^{1/4}T_j|\,|\delta^{-1/4}M_j| \,|\, \overline{X}, \alpha) + \tfrac{1}{2}t^2\,\tilde\alpha_j\,\varphi_j^2 \\
&\le 1 - \tfrac{1}{2}t^2\,\tilde\alpha_j\,(s_j^2 - \tfrac{1}{3}\beta_j\,\beta^{-1} - 2(\delta^{1/2}\,s_j^2)^{1/2}\,(\delta^{-1/2}\,\varphi_j^2)^{1/2} - \varphi_j^2) \\
&\le \exp\{-\tfrac{1}{2}t^2\,\tilde\alpha_j\,((1-\delta^{1/2})\,s_j^2 - \tfrac{1}{3}\beta_j\,\beta^{-1} - (1+\delta^{-1/2})\,\varphi_j^2)\}. \qquad (2.38)
\end{aligned}$$

Here we used Hölder's inequality, the fact that in general both $2\,x^{1/2}\,y^{1/2} \le x + y$ (for $x, y \ge 0$) and $1 - x \le \exp\{-x\}$, and the fact that by construction of $\tilde\alpha_j$ we have that $1 - \frac{1}{2}t^2\,\tilde\alpha_j\,s_j^2 \ge 0$. As to the latter: in case $\tilde\alpha_j = 1$ it is clear that $1 \ge \frac{1}{2}\beta^{-2}\,s_j^2$, whereas in case $1 < \frac{1}{2}\beta^{-2}\,s_j^2$ it is clear that $s_j^2 < (\frac{1}{2}\beta^{-2}\,s_j^2)^{1/2}\,s_j^2 \le 2^{-1/2}\,\beta_j\,\beta^{-1}$ and therefore $\tilde\alpha_j = 0$.

Writing $r_j := (1-\delta^{1/2})\,s_j^2 - \frac{1}{3}\beta_j\,\beta^{-1}$, we see that

$$\begin{aligned}
I\{\varphi\}\,\textstyle\prod_{j=1}^n W_j &\le I\{\varphi\}\,\exp\{-\tfrac{1}{2}t^2\,\textstyle\sum_{j=1}^n \tilde\alpha_j\,(r_j - (1+\delta^{-1/2})\,\varphi_j^2)\} \\
&\le I\{\varphi\}\,\exp\{\tfrac{1}{2}t^2\,\textstyle\sum_{j=1}^n \tilde\alpha_j\,(1+\delta^{-1/2})\,\varphi_j^2\}\,\exp\{-\tfrac{1}{2}t^2\,\textstyle\sum_{j=1}^n \tilde\alpha_j\,r_j\}. \qquad (2.39)
\end{aligned}$$

Now we have that

$$\begin{aligned}
\varepsilon_{23} &\le \textstyle\int_{a^p}^{a^q} \exp\{\tfrac{1}{2}m\,t^2\,\delta\,(1+\delta^{-1/2})\}\,\mathbb{E}\,\exp\{-\tfrac{1}{2}t^2\,\textstyle\sum_{j=1}^n \tilde\alpha_j\,r_j\}\,dt \\
&= \textstyle\int_{a^p}^{a^q} \exp\{\tfrac{1}{2}m\,t^2\,\delta^{1/2}\,(1+\delta^{1/2})\}\,\textstyle\prod_{j=1}^n \mathbb{E}\,\exp\{-\tfrac{1}{2}t^2\,\tilde\alpha_j\,r_j\}\,dt. \qquad (2.40)
\end{aligned}$$

Taking $Z, Z_1, \ldots, Z_n$ to be an independent sample from the standard normal distribution, we can write

$$v_j \;:=\; \mathbb{E}\,\exp\{-\tfrac{1}{2}t^2\,\tilde\alpha_j\,r_j\} \;=\; \mathbb{E}\,e\{t\,\alpha_j\,(\gamma_j\,r_j)^{1/2}\,Z_j\}.$$

Note that $\gamma_j\,r_j \ge 0$ for $j = 1, \ldots, n$ (that is to say: $r_j < 0 \Rightarrow \gamma_j = 0$) if we take $\delta$ small enough, which we will. Expansion into Taylor series leads to the following:

$$\begin{aligned}
v_j &\le 1 - \tfrac{1}{2}m\,t^2\,\gamma_j\,r_j + \tfrac{1}{6}\,m\,t^3\,\mathbb{E}\,|Z|^3\,(\gamma_j\,r_j)^{3/2} \\
&\le 1 - \tfrac{1}{2}m\,t^2\,\gamma_j\,(r_j - \tfrac{1}{3}\mathbb{E}\,|Z|^3\,(1-\delta^{1/2})^{3/2}\,s_j^3\,\beta^{-1}) \\
&\le \exp\{-\tfrac{1}{2}m\,t^2\,\gamma_j\,w_j\}, \qquad (2.41)
\end{aligned}$$

with

$$w_j \;:=\; (1-\delta^{1/2})\, s_j^2 - \tfrac{1}{3}(1 + 4(2\pi)^{-1/2}\,(1-\delta^{1/2})^{3/2})\,\beta_j\,\beta^{-1},$$

and as a result

$$\varepsilon_{23} \;\le\; \int_{a^p}^{a^q} \exp\{-\tfrac{1}{2} m\, t^2\,(\textstyle\sum_{j=1}^n \gamma_j w_j - \delta^{1/2}\,(1+\delta^{1/2}))\}\,dt.$$

Finally we note that $\sum_{j=1}^n \gamma_j w_j \ge \sum_{j=1}^n w_j$ because $\gamma_j = 0 \;\Rightarrow\; w_j \le 0$ for $\delta$ small enough, so that

$$\begin{aligned}\varepsilon_{23} \;&\le\; \int_{a^p}^{a^q} \exp\{-\tfrac{1}{2} m\, t^2\,(\textstyle\sum_{j=1}^n w_j - \delta^{1/2}\,(1+\delta^{1/2}))\}\,dt \\ &=\; \int_{a^p}^{a^q} \exp\{-\tfrac{1}{2}\tilde{\delta}\, m\, t^2\}\,dt,\end{aligned} \tag{2.42}$$

with

$$\tilde{\delta} \;:=\; 1 - 2\,\delta^{1/2} - \tfrac{1}{3} - \tfrac{4}{3}\,(2\pi)^{-1/2}\,(1-\delta^{1/2})^{3/2} - \delta. \tag{2.43}$$

Taking $\delta > 0$ such that $r_j < 0 \;\Rightarrow\; \gamma_j = 0$, $\gamma_j = 0 \;\Rightarrow\; w_j \le 0$ and $\tilde{\delta} > 0$, we see as in (2.20) that $\varepsilon_{23} \ll_{p,q} 1$. Thus $\varepsilon_{21} \ll_{p,q} 1$ and $\varepsilon_2 \ll_{p,q} 1$, which completes the proof. □

## 2.5 Proof of Theorem 2.2 for arbitrary statistics

We go about in the same way as for $U$-statistics of order 2. Again we may assume without loss of generality that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We put

$$a \;:=\; \min\{\beta^{-1},\; \Delta^{-2}\}, \qquad a^{-1} \;=\; \max\{\beta,\; \Delta^2\}$$

and prove that $\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\,dt$ is bounded by some absolute constant. We may assume that $a \ge 100$. We set

$$\varepsilon_1 \;:=\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}\}|\,dt \quad \text{and} \quad \varepsilon_2 \;=\; \varepsilon_2(p,q) \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\,dt$$

with $\frac{1}{4} < q \le 1$ and $q - \frac{1}{4} < p < q$ and prove that $\varepsilon_1 \ll 1$ and $\varepsilon_2 \ll_{p,q} 1$, after which Theorem 2.2 is an obvious consequence.

We start with $\varepsilon_1$. Here we have that

$$\varepsilon_1 \;\le\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}_1\}|\,dt + \int_0^{a^{1/4}} t\,\mathbb{E}\,|\mathbb{T}_2 + \mathbf{T}_3|\,dt$$

with (see (A.17))

$$\int_0^{a^{1/4}} t\,\mathbb{E}\,|\mathbb{T}_2+\mathbf{T}_3|\,dt \;\le\; \tfrac{1}{2}\,\{\mathbb{E}\,|\mathbb{T}_2+\mathbf{T}_3|^2\}^{1/2}\,a^{1/2} \;\le\; \tfrac{1}{2}(\Delta^2)^{1/2}\,a^{1/2} \;\le\; \tfrac{1}{2},$$

and moreover, taking $\tilde{a} := \min\{\beta^{-1}, \Delta_1^{-2}\}$, by the proof of Lemma 2.6,

$$\int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}_1\}|\,dt \;\le\; \int_0^{\tilde{a}^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}_1\}|\,dt \;\ll\; 1:$$

as a consequence $\varepsilon_1 \ll 1$.

Using the method which was used in the proof of Lemma 2.7 we show that

$$\varepsilon_2 \;\ll_{p,q}\; 1.$$

To this end we take again an independent copy $\overline{X}_1, \dots, \overline{X}_n$ of $X_1, \dots, X_n$, together with an i.i.d. sample $\alpha_1, \dots, \alpha_n$ with

$$\mathbb{P}(\alpha_1 = 1) \;=\; 1 - \mathbb{P}(\alpha_1 = 0) \;=\; m \;:=\; a^{-(2q-\frac{1}{2})}.$$

Now let for any pair $A, B \subset N$ with $A \cap B = \emptyset$

$$T_{A,\overline{B}} \;:=\; T_{A,B}(\{X_j : j \in A\}, \{\overline{X}_k : k \in B\})$$

(that is to say, by $T_{1,4,\overline{2},\overline{3}}$ in fact we mean $T_{1,2,3,4}(X_1, \overline{X}_2, \overline{X}_3, X_4)$, and so on), and let for any subset $A = \{n_1, \dots, n_k\} \subset N$

$$\alpha_A \;:=\; \prod_{j=1}^k \alpha_{n_j} \quad \text{and} \quad (1-\alpha)_A \;:=\; \prod_{j=1}^k (1-\alpha_{n_j}).$$

We look at $\tilde{\mathbb{T}}$ instead of $\mathbb{T}$, where

$$\tilde{\mathbb{T}} \;:=\; \tilde{\mathbb{T}}_{11} + \sum_{j=1}^n \alpha_j\,(\tilde{M}_j + \tilde{N}_j) + \tilde{\Lambda} + \tilde{R}, \tag{2.44}$$

with $\tilde{\mathbb{T}}_{11}$ as in (2.31),

$$\tilde{M}_j \;:=\; \sum_{k:k\ne j}(1-\alpha_k)\,T_{j,\overline{k}}, \quad \tilde{N}_j \;:=\; \sum_{B:\,j\notin B,|B|\ge 2}(1-\alpha)_B\,T_{j,\overline{B}},$$

for $j = 1, \dots, n$,

$$\tilde{\Lambda} \;:=\; \sum_{A:\,|A|\ge 2} \alpha_A \sum_{B:\,A\cap B=\emptyset} (1-\alpha)_B\,T_{A,\overline{B}},$$

and $\tilde{R} := \sum_B (1-\alpha)_B\,T_{\overline{B}}$. This decomposition is the analogue to the one which can be found in (2.23).

Analogue to Lemma 2.5, we need bounds on the moments of $\tilde{\Lambda}$, the $\tilde{M}_j$ and the $\tilde{N}_j$. The bounds are much similar to the ones obtained before, and read as follows:

**Lemma 2.8.** *We have:*

$$\mathbb{E}\,\tilde{\Lambda}^2 \le m^2\,\Delta^2, \quad \textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{M}_j^2 \le 2\,\Delta^2, \quad \textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{N}_j^2 \le \Delta^2.$$

**Proof of Lemma 2.8.** Going about as in (A.18), we see that

$$\begin{aligned}\mathbb{E}\,\tilde{\Lambda}^2 &= \textstyle\sum_{A:\,|A|\ge 2} m^{|A|} \sum_{B:\,A\cap B=\emptyset} (1-m)^{|B|}\,\mathbb{E}\,T^2_{A,\overline{B}} \\ &= \textstyle\sum_{l=2}^n \sum_{C:\,|C|=l} \left\{ \sum_{k=2}^l \binom{l}{k} m^k\,(1-m)^{l-k} \right\} \mathbb{E}\,T_C^2. \qquad (2.45)\end{aligned}$$

Indeed, this corresponds to the fact that, given any subset $C \subset N$ with $|C| = l \ge 2$, restricting ourselves to sets $A$ for which $|A| = k$, $2 \le k \le l$, we come $\binom{l}{k}$ times across $\mathbb{E}\,T_C^2$. We notice that

$$\begin{aligned}&\textstyle\sum_{k=2}^l \binom{l}{k} m^k\,(1-m)^{l-k} \\ &\quad= m^2\,l(l-1) \textstyle\sum_{k=2}^l \frac{1}{k(k-1)} \frac{(l-2)!}{(k-2)!\,(l-k)!}\, m^{k-2}\,(1-m)^{l-k} \\ &\quad\le m^2 \binom{l}{2} \textstyle\sum_{r=0}^{l-2} \binom{l-2}{r} m^r\,(1-m)^{(l-2)-r} \;=\; m^2\binom{l}{2},\end{aligned}$$

using the substitution $r := k-2$ and Newton's binomial formula, and from (2.45) and (A.18) it is clear that

$$\mathbb{E}\,\tilde{\Lambda}^2 \;\le\; m^2\,\Delta^2.$$

As to the $\tilde{M}_j$, we see that

$$\textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{M}_j^2 \;\le\; \sum_{j=1}^n \sum_{k:\,k\ne j} \mathbb{E}\,T^2_{j,\overline{k}} \;=\; 2\,\Delta_1^2 \;\le\; 2\,\Delta^2.$$

Finally, using the fact that $l \le \binom{l}{2}$ for $l = 3, \ldots, n$, we see that

$$\begin{aligned}\textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{N}_j^2 &\le \textstyle\sum_{j=1}^n \sum_{B:\,j\notin B,|B|\ge 2} \mathbb{E}\,T^2_{j,\overline{B}} \\ &= \textstyle\sum_{l=3}^n \sum_{C:\,|C|=l} l\,\mathbb{E}\,T_C^2 \;\le\; \Delta^2,\end{aligned}$$

again using (A.18). This finishes the proof. □

We continue with our estimation of

$$\varepsilon_2 \;=\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\,e\{t\tilde{\mathbb{T}}\}|\,dt.$$

First we remove $\tilde{\Lambda}$ by means of a simple Taylor expansion. This leads us to the following:

$$\varepsilon_2 \;\le\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t(\tilde{\mathbb{T}} - \tilde{\Lambda})\}|\, dt + \int_{a^p}^{a^q} t\, \mathbb{E}\, |\tilde{\Lambda}|\, dt \;=:\; \varepsilon_{21} + \varepsilon_{22}.$$

As to $\varepsilon_{22}$ we have that $\mathbb{E}\tilde{\Lambda}^2 \le m^2\, \Delta^2 \le m^2\, a^{-1}$, so that

$$\varepsilon_{22} \;\le\; \{m^2\, a^{-1}\}^{1/2}\, \tfrac{1}{2} a^{2q} \;\le\; \tfrac{1}{2} m\, a^{2q-\frac{1}{2}} \;=\; \tfrac{1}{2}.$$

This leaves us only the estimation of

$$\varepsilon_{21} \;=\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\, (\tilde{\mathbb{T}}_{11} + \textstyle\sum_{j=1}^n \alpha_j\, (\tilde{M}_j + \tilde{N}_j) + \tilde{R})\}|\, dt.$$

We proceed as with the $U$-statistics of order 2.

Taking

$$V_j \;:=\; |\mathbb{E}\, (e\{t\, \alpha_j (T_j + \tilde{M}_j + \tilde{N}_j)\} \,|\, \overline{X}, \alpha)|,$$

we have (2.34), thus removing $\tilde{R}$ from our considerations as well. We go about in the exact same way as in Section 2.4, the only difference being that we take

$$\varphi_j^2 \;:=\; \mathbb{E}\, ((\tilde{M}_j + \tilde{N}_j)^2 \,|\, \overline{X}, \alpha) \;=\; \mathbb{E}\, (\tilde{M}_j^2 + \tilde{N}_j^2 \,|\, \overline{X}, \alpha) \tag{2.46}$$

instead of (2.33) (note that, conditionally on $\overline{X}$ and $\alpha$, $\tilde{M}_j$ and $\tilde{N}_j$ are uncorrelated). Again we use the indicator $I\{\varphi\}$ as in (2.35) to ensure that $\sum_{j=1}^n \tilde{\alpha}_j\, \varphi_j^2 \le \delta\, m$. The cost of getting this indicator in is bounded by

$$\begin{aligned} \mathbb{P}(\textstyle\sum_{j=1}^n \tilde{\alpha}_j\, \varphi_j^2 > \delta\, m) \;&\le\; \delta^{-1}\, m^{-1}\, m\, \textstyle\sum_{j=1}^n (\mathbb{E}\, \tilde{M}_j^2 + \mathbb{E}\, \tilde{N}_j^2) \\ &\le\; \delta^{-1}\, (2\Delta^2 + \Delta^2) \;\le\; 3\delta^{-1}\, a^{-1} \end{aligned} \tag{2.47}$$

(see both (2.37) and Lemma 2.8). After this the fact that $\varepsilon_2 \ll_{p,q} 1$ is proved in exactly the same way as in Section 2.4. This concludes the proof of Theorem 2.2.

# Chapter 3

# Berry-Esseen bounds

## 3.1 Introduction and results

Let $X_1, \ldots, X_n$ be independent, not necessarily identically distributed random variables, taking their values in arbitrary measurable spaces $(\mathcal{X}_j, \mathcal{B}_j)$. As earlier, we consider arbitrary, real-valued statistics

$$\mathbb{T} \;=\; \mathbb{T}(X_1, \ldots, X_n),$$

for which $\mathbb{E}|\mathbb{T}| < \infty$. Let $\sigma^2 = \mathrm{var}(\mathbb{T})$ and $s^2 = \mathrm{var}(\mathbb{T}_1)$ as in (A.14). We assume that $0 < s^2 < \infty$ and shall denote

$$\tilde{\mathbb{T}} \;:=\; (\mathbb{T} - \mathbb{E}\,\mathbb{T})/s.$$

Let $\Phi$ be the distribution function of the standard normal distribution, and let

$$D \;=\; D(\mathbb{T}) \;:=\; \sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\mathbb{T}} \le x) - \Phi(x)|. \tag{3.1}$$

By a *Berry-Esseen bound* for the statistic $\mathbb{T}$ we mean a bound for (3.1), which, under appropriate conditions on the sequence of statistics and the samples, is expected to satisfy $D(\mathbb{T}) = \mathcal{O}(n^{-1/2})$ (as $n \to \infty$).

Looking at statistics of the form

$$\mathbb{T} \;=\; \sum_{1\le j_1 < \ldots < j_k \le n} h(X_{j_1}, \ldots, X_{j_k}),$$

for some $k \geq 2$, symmetric kernel $h$ and i.i.d. sample, Berry-Esseen bounds for these statistic were considered by, e.g., Chan and Wierman (1977), Callaert and Janssen (1978) and Helmers and Van Zwet (1983), who proved in increasing generality (that is to say, posing ever lighter conditions on the underlying distribution), that indeed $D = \mathcal{O}(n^{-1/2})$. Van Zwet (1984) covers the more general situation in which $\mathbb{T}$ is an arbitrary symmetric function of an i.i.d. sample. Using the notation $\tilde{\beta}_3$ and $\tilde{\Delta}_2^2$ as in (A.20), in this paper the existence of an absolute constant $c$ is derived such that

$$D(\mathbb{T}) \;\leq\; c\left(\tilde{\beta}_3\, s^{-3} + \tilde{\Delta}_2^2\, s^{-2}\right)\, n^{-1/2}, \tag{3.2}$$

thus generalizing the earlier results. As applications, nice results are obtained for both $U$-statistics and linear combinations of order statistics.

As concerns the power of $n$, the expression $\tilde{\beta}_3\, s^{-3}\, n^{-1/2}$ in the bound cannot be improved. Indeed, if we take an i.i.d. sample $X_1, \ldots, X_n$ such that

$$\mathbb{P}(X_j = 1) \;=\; \mathbb{P}(X_j = -1) \;=\; \tfrac{1}{2},$$

and take as our statistic $\mathbb{T} = \sum_{j=1}^n X_j$, an absolute constant $c_1$ is easily found such that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\mathbb{T}} \leq x) - \Phi(x)| \;\geq\; c_1\, \tilde{\beta}_3\, s^{-3}\, n^{-1/2}.$$

The same example shows that taking higher order moments will not help us in this respect (since always $\tilde{\beta}_p\, s^{-p} = 1$).

We start from the following two intuitive notions, to which our results are to answer:

1. Any i.i.d. result must be easily translatable to a non-i.i.d. analogue,

2. Any non-i.i.d. result must not explicitly depend on $n$.

Bearing this in mind, we contemplate on the bound (3.2). Taking $\beta$ as in (A.14), we already noticed that $\beta = \tilde{\beta}_3\, n^{-1/2}$. Thinking of our earlier results on concentration functions, the i.i.d. expression $\tilde{\beta}_3\, n^{-1/2}$ is thus easily generalized to the non-i.i.d. expression $\beta$. On the other hand, (A.24) is telling us that $\Delta^2 \approx \frac{1}{2}\tilde{\Delta}_2^2\, n^{-1}$, so the expression $(\tilde{\Delta}_2^2\, s^{-2}\, n^{-1})^{1/2}$ would make more sense in the bound (3.2) than the actual $\tilde{\Delta}_2^2\, s^{-2}\, n^{-1/2}$.

Now let $\tilde{\gamma}_2$ and $\tilde{\delta}_2$ be as defined in (A.20) and (A.21), and write

$$\kappa \;:=\; n^{5/2}\, \mathbb{E}\, T_1 T_2\, T_{1,2}.$$

In the i.i.d. symmetric case we obtain the following Berry-Esseen type result. Let

$$G(x) \;:=\; \Phi(x) + \tfrac{1}{2} n^{-1/2}\,\kappa\, s^{-3}\,\Phi'''(x) \tag{3.3}$$

denote an asymptotic expansion belonging to $\mathbb{T}$. We have the following:

**Theorem 3.1.** *Assume that the sample $X_1, \dots, X_n$ is i.i.d., the statistic $\mathbb{T}$ is symmetric, and $0 < s^2 < \infty$. Then there exists a constant $c$ such that*

$$D(\mathbb{T}) \;\le\; c\,\max\{\tilde\beta_3\, s^{-3}\, n^{-1/2},\ (\tilde\Delta_2^2\, s^{-2}\, n^{-1})^{1/2}\}. \tag{3.4}$$

*Furthermore, for any $0 < \varepsilon < 1$, a constant $c(\varepsilon)$ exists such that*

$$\begin{aligned} &\sup\nolimits_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\mathbb{T}} \le x) - G(x)| \\ &\quad\le\; c(\varepsilon)\,\max\{\tilde\beta_3\, s^{-3}\, n^{-1/2},\ (\tilde\gamma_2\, s^{-2}\, n^{-1})^{1-\varepsilon},\ (\tilde\delta_2\, s^{-2}\, n^{-1})^{1/2}\}. \end{aligned} \tag{3.5}$$

*In the case that moreover $\sigma^2 < \infty$, we can replace $s$ in (3.4) and (3.5) by $\sigma$.*

Notice that our Berry-Esseen bound (3.4) improves (3.2). Indeed, by Lyapunov's inequality, $\tilde\beta_3\, s^{-3} = \mathbb{E}\,|T_1|^3/(\mathbb{E}\,T_1^2)^{3/2} \ge 1$. As a result, taking

$$\delta \;:=\; \max\{\tilde\beta_3\, s^{-3}\, n^{-1/2},\ (\tilde\Delta_2^2\, s^{-2}\, n^{-1})^{1/2}\},$$

in the case that $\tilde\Delta_2^2\, s^{-2} \le 1$ we have $\delta = \tilde\beta_3\, s^{-3}\, n^{-1/2}$, whereas otherwise

$$\delta \;\le\; \max\{\tilde\beta_3\, s^{-3}\, n^{-1/2},\ \tilde\Delta_2^2\, s^{-2}\, n^{-1/2}\},$$

and (3.4) easily leads to (3.2).

The asymptotic expansion $G(x)$ is kept into our result because it is sometimes, e.g. for self-normalized statistics, possible to derive better Berry-Esseen bounds from it than can be obtained by just bluntly using the general bound

$$\|G - \Phi\|_\infty \;\le\; \tfrac{1}{2}(\tilde\gamma_2\, s^{-2}\, n^{-1})^{1/2}\, \|\Phi'''\|_\infty, \tag{3.6}$$

which follows from (3.3) and the fact that, by Hölder's inequality,

$$|\kappa|\, s^{-3} \;\le\; \{\mathbb{E}\,|n\, T_1 T_2|^2\}^{1/2}\, \{\mathbb{E}\,|n^{3/2} T_{1,2}|^2\}^{1/2}\, s^{-3} \;=\; (\tilde\gamma_2\, s^{-2})^{1/2}.$$

Note that (3.4) is an immediate consequence of (3.5) because of (3.6).

The proof of (3.5) is much similar to that of Theorem 2.1. We start by giving the proof under the assumption that $\mathbb{T}$ is a $U$-statistic of order 2, and then later on extend the proof to the case of arbitrary statistics.

Instead of $(\tilde\gamma_2\, s^{-2}\, n^{-1})^{1-\varepsilon}$, we may be looking for an expression in the bound that contains the term $n^{-1}$. Making higher moment assumptions, we may well obtain this kind of bound. We have the following extension to Theorem 3.1:

**Corollary 3.2.** *The term* $(\tilde{\gamma}_2\, s^{-2}\, n^{-1})^{1-\varepsilon}$ *in* (3.5) *may be replaced by*

$$\tilde{\gamma}_{2+\varepsilon}\, s^{-(2+\varepsilon)}\, n^{-1} \quad or \quad (\tilde{\gamma}_2\, s^{-2})^{1+\varepsilon}\, n^{-1}.$$

The proof of Corollary 3.2 will be given in Section 3.3, after the proof of Theorem 3.1.

We turn to the more general case where the sample $X_1, \ldots, X_n$ is not necessarily identically distributed, and where the statistic $\mathbb{T}$ is not necessarily symmetric in its arguments. This situation has been considered by Friedrich (1989), who obtains the following bound:

$$D(\mathbb{T}) \;\leq\; c(p)\,\big(\beta\, s^{-3} + n^{3/2}\,\gamma_1\,\gamma_{2,p} + n^2\,\gamma_1^2\,\gamma_{3,3/2}\big), \tag{3.7}$$

for any $p \in [\frac{3}{2}, 2)$,

$$\gamma_1 \;:=\; s^{-1}\,\max\nolimits_{1\leq j\leq n}\beta_j^{1/3}, \qquad \gamma_{2,p} \;:=\; s^{-p}\,\max\nolimits_{1\leq j\leq n}\mathbb{E}\,|D_j\mathbb{T} - T_j|^p,$$

and

$$\gamma_{3,p} \;:=\; s^{-1}\,\big(\max\nolimits_{1\leq j<k\leq n}\mathbb{E}\,|D_{j,k}\mathbb{T}|^p\big)^{1/p}.$$

Under appropriate moment conditions this gives the desired order of $n^{-1/2}$, and as applications appropriate Berry-Esseen theorems are derived for $U$-statistics, rank statistics and linear combinations of order statistics, starting from non-i.i.d. samples. A weak point of the bound (3.7) is that it uses maxima of moments instead of sums. Moreover, the multiplicative structure is not too transparant. In this respect, the following is more satisfactory. Let the moments $\Delta^2$, $\Delta_1^2$ and $\Delta_2^2$ be as defined in (A.15) and (A.16), and let

$$H(x) \;:=\; \Phi(x) + s^{-3}\,\big(\textstyle\sum_{1\leq j<k\leq n}\mathbb{E}\,T_j T_k T_{j,k}\big)\,\Phi'''(x)$$

be the non-i.i.d. analogue to $G$. We have the following:

**Theorem 3.3.** *Let* $\varepsilon \in (0,1)$ *be fixed. Suppose that the sample is independent but not necessarily identically distributed, and that* $0 < s^2 < \infty$. *Then there exists a constant* $c(\varepsilon)$ *such that*

$$D(\mathbb{T}) \;\leq\; c(\varepsilon)\,\max\big\{\beta\, s^{-3},\; (\Delta_1^2\, s^{-2})^{1/2},\; (\Delta_2^2\, s^{-2})^{(1-\varepsilon)/2}\big\}\,. \tag{3.8}$$

*Furthermore, for any* $0 < \varepsilon < 1$ *a constant* $c(\varepsilon)$ *exists such that*

$$\begin{aligned} &\sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\mathbb{T}} \le x) - H(x)| \\ &\quad\le\; c(\varepsilon)\, \max\big\{\beta\, s^{-3},\ (\Delta_1^2\, s^{-2})^{1-\varepsilon},\ (\Delta_2^2\, s^{-2})^{(1-\varepsilon)/2}\big\}\,. \end{aligned} \tag{3.9}$$

*In the case that moreover* $\sigma^2 < \infty$*, we can replace* $s$ *in* (3.8) *and* (3.9) *by* $\sigma$*.*

Note that the Berry-Esseen bound (3.8) is an easy consequence of (3.9). Indeed, with Hölder's inequality and the Cauchy-Schwarz inequality

$$\begin{aligned} \textstyle\sum_{1\le j<k\le n} \mathbb{E}\,|T_j T_k T_{j,k}| \;&\le\; \textstyle\sum_{1\le j<k\le n} \{\mathbb{E}\,|T_j T_k|^2\}^{1/2}\,\{\mathbb{E}\,T_{j,k}^2\}^{1/2} \\ &\le\; \{\textstyle\sum_{1\le j<k\le n} s_j^2\, s_k^2\}^{1/2}\,\{\textstyle\sum_{1\le j<k\le n} \mathbb{E}\,T_{j,k}^2\}^{1/2} \\ &\le\; (\textstyle\sum_{j=1}^n s_j^2)\,(\Delta_1^2)^{1/2} \;=\; s^3\,(\Delta_1^2\, s^{-2})^{1/2}, \end{aligned} \tag{3.10}$$

which leads us to the fact that

$$\|H-\Phi\|_\infty \;\le\; \tfrac{1}{2}\,\|\Phi'''\|_\infty\,(\Delta_1^2\, s^{-2})^{1/2},$$

after which then (3.9) leads us to the desired result.

The proof of (3.9) will be given in the Sections 3.4 and 3.5. At the end of Section 3.5 we show that, apart from the $\varepsilon$ in the power of $\Delta_2^2\, s^{-2}$, Theorem 3.3 is a generalization of Theorem 3.1.

## 3.2 Proof of Theorem 3.1 for $U$-statistics of order 2

The basis for the proof of the theorems is the following approximate Fourier inversion formula:

$$K(x) \;=\; \frac{1}{2} + \frac{i}{2\pi}\int_{-a}^{a} e\{-tx\}\,\hat{k}(t)\,\frac{dt}{t} + R, \tag{3.11}$$

with

$$|R| \;\le\; a^{-1}\textstyle\int_{-a}^{a} |\hat{k}(t)|\,dt,$$

see for example Prawitz (1972), or Bentkus and Götze (1996b), Lemma 4.1. Here we assume that $a > 0$ is an arbitrary fixed positive number, that $K$ is a function of bounded variation with

$$\lim_{x\to-\infty} K(x) = 0 \quad\text{and}\quad \lim_{x\to\infty} K(x) = 1,$$

and that $\hat{k}(t) := \int e\{tx\}\, dK(x)$ is the Fourier transform of $K$.

Let from now on $\hat{g}$ be the Fourier transform of $G$ and

$$\hat{f}(t) \;:=\; \mathbb{E}\, e\{t\mathbb{T}\}.$$

In the present case, it follows from (3.11) that, for all $a > 0$ and $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(\mathbb{T} \leq x) - G(x) \;&=\; \tfrac{i}{2\pi} \int_{-a}^{a} e\{-tx\} \left(\hat{f}(t) - \hat{g}(t)\right) t^{-1}\, dt + R_1 - R_2 \\ &=\; \tfrac{i}{\pi}\, \mathrm{Re} \int_0^a e\{-tx\} \left(\hat{f}(t) - \hat{g}(t)\right) t^{-1}\, dt + R_1 - R_2,\end{aligned}$$

and hence

$$|\mathbb{P}(\mathbb{T} \leq x) - G(x)| \;\leq\; \pi^{-1} \int_0^a |\hat{f}(t) - \hat{g}(t)|\, t^{-1}\, dt + |R_1| + |R_2|, \tag{3.12}$$

with

$$|R_1| \;\leq\; a^{-1} \int_{-a}^{a} |\hat{f}(t)|\, dt \quad \text{and} \quad |R_2| \;\leq\; a^{-1} \int_{-a}^{a} |\hat{g}(t)|\, dt.$$

This will be our point of departure.

We start by proving Theorem 3.1 in the case that $\mathbb{T}$ is a $U$-statistic of order 2, that is to say, of the form (A.6) with $\mathbf{T}_3 = 0$, so that $\tilde{\delta}_2 = 0$. Like before we may assume that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We take

$$a \;:=\; \min\{\tilde{\beta}_3^{-1}\, n^{1/2},\; (\tilde{\gamma}_2^{-1}\, n)^{1-\varepsilon}\} \;\geq\; 100 \tag{3.13}$$

and are going to apply (3.12). Notice that in general

$$\tilde{\gamma}_2\, n^{-1} \;\leq\; (a^{-1})^{1/(1-\varepsilon)} \;=\; a^{-k/(k-1)} \;=\; a^{-1-1/(k-1)}, \tag{3.14}$$

and that furthermore, as in (2.5),

$$\tilde{\beta}_3 \;\geq\; 1 \qquad \text{and} \qquad n \;\geq\; a^2 \;\geq\; 100^2. \tag{3.15}$$

We assume as well that $\varepsilon = k^{-1}$ for some natural number $k \geq 2$. Indeed, otherwise we just prove the desired result with $\varepsilon^* := ([\varepsilon^{-1}]+1)^{-1} < \varepsilon$ instead of $\varepsilon$, which is stronger than the result we are looking for. We use the notation $b_1 \ll b_2$ as in (2.1), and start by removing the remainder terms $R_1$ and $R_2$.

As to $R_1$: we proved already that $\int_{-\tilde{a}}^{\tilde{a}} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \ll 1$, see (2.6), with $\tilde{a} = \min\{\tilde{\beta}_3^{-1}\, n^{1/2},\; \tilde{\gamma}_2^{-1}\, n\} \geq a$, so indeed

$$|R_1| \;\ll\; a^{-1}.$$

As to $R_2$: writing

$$k_n(t) \;:=\; \tfrac{1}{2}(it)^3\, n^{-1/2}\,\kappa \qquad \text{and} \qquad \tilde{k}_n(t) \;:=\; 1 + k_n(t)$$

we have that $\hat{g}(t) = \tilde{k}_n(t)\,\exp\{-\frac{1}{2}t^2\}$, and setting

$$l_n(t) \;:=\; \tfrac{1}{2}\,|t|^3\,(\tilde{\gamma}_2\, n^{-1})^{1/2} \qquad \text{and} \qquad \tilde{l}_n(t) \;:=\; 1 + l_n(t),$$

moreover $|\tilde{k}_n(t)| \le \tilde{l}_n(t)$ for all $t$, which shows us that, cf. (3.13),

$$\textstyle\int_{-a}^{a} |\hat{g}(t)|\, dt \;\le\; 2\,\int_0^a \tilde{l}_n(t)\,\exp\{-\frac{1}{2}t^2\}\, dt \;\ll\; 1.$$

As a consequence we have as well that $|R_2| \ll a^{-1}$, which completes the removal of $R_1$ and $R_2$.

From here on we shall use the notation $f_1 \sim f_2$, meaning by definition that

$$\sup\nolimits_{x\in\mathbb{R}} \left|\textstyle\int_0^a e\{-tx\}\;(f_1(t) - f_2(t))\;t^{-1}\,dt\right| \;\ll\; a^{-1},$$

where, for $j = 1, 2$, we assume that $f_j$ is a complex-valued function for which $\int_0^a |f_j(t)|\,t^{-1}\,dt$ exists. Clearly $\sim$ defines an equivalence relation on the class of these functions, and we have to show that $\hat{f} \sim \hat{g}$.

Now let $\mathbb{T}_1$ as in (A.7) denote the linear part of $\mathbb{T}$, and write

$$\hat{h}_1(t) := \tilde{k}_n(t)\,\mathbb{E}\,e\{t\mathbb{T}_1\}, \quad \hat{h}_2(t) := \mathbb{E}\,e\{t\mathbb{T}_1\} + k_n(t)\,\mathbb{E}\,e\{t(\mathbb{T}_1 - (T_1 + T_2))\}.$$

The following lemma reflects that for our purposes it makes no difference whether we look at $\hat{g}$, $\hat{h}_1$ or $\hat{h}_2$.

**Lemma 3.4.** *We have that $\hat{g} \sim \hat{h}_1$ and that $\hat{h}_1 \sim \hat{h}_2$.*

**Proof of Lemma 3.4.** Let $Z_1, \ldots, Z_n$ be an i.i.d. sample of random variables that are $N(0, \frac{1}{n})$-distributed, independent of the sample. Writing $Z := \sum_{j=1}^n Z_j$, we have that $Z$ is standard normally distributed. Now take any $x \in \mathbb{R}$ and set

$$\delta_1 \;:=\; \left|\textstyle\int_0^a e\{-tx\}\;\Big(\hat{g}(t) - \hat{h}_1(t)\Big)\;t^{-1}\,dt\right|.$$

We will show that $\delta_1 \ll a^{-1}$ so that $\hat{g} \sim \hat{h}_1$.

Indeed, we have that

$$\delta_1 \;\le\; \int_0^a |\hat{g}(t) - \hat{h}_1(t)|\, t^{-1}\, dt \;\le\; \int_0^a \tilde{l}_n(t)\, |\mathbb{E}\, e\{t\mathbb{T}_1\} - \mathbb{E}\, e\{tZ\}|\; t^{-1}\, dt.$$

In order to estimate $|\mathbb{E}\, e\{t\mathbb{T}_1\} - \mathbb{E}\, e\{tZ\}|$ we proceed in small steps. Looking at the Taylor expansion (see (2.8)) of $e\{tT_j\}$ around 0, we see that

$$\mathbb{E}\, e\{tT_j\} \;=\; 1 - \tfrac{1}{2} n^{-1}\, t^2 + \tfrac{1}{2}(it)^3\, \mathbb{E}\,(1-\tau)^2 e\{t\tau T_j\}\, T_j^3, \tag{3.16}$$

whereas

$$\mathbb{E}\, e\{tZ_j\} \;=\; 1 - \tfrac{1}{2} n^{-1}\, t^2 + \tfrac{1}{2}(it)^3\, \mathbb{E}\,(1-\tau)^2 e\{t\tau Z_j\}\, Z_j^3. \tag{3.17}$$

Notice that as well $\mathbb{E}\, e\{tZ_j\} = \exp\{-\frac{1}{2} n^{-1}\, t^2\}$ and that for $0 \le t \le a$ we have that

$$\begin{aligned}
|\mathbb{E}\, e\{tT_j\}| \;&\le\; |1 - \tfrac{1}{2} n^{-1}\, t^2| + \tfrac{1}{6} t^3\, \mathbb{E}\,|T_1|^3 \\
&\le\; 1 - \tfrac{1}{2} n^{-1}\, t^2 + \tfrac{1}{6} \tilde{\beta}_3^{-1}\, n^{1/2}\, n^{-3/2}\, \tilde{\beta}_3\, t^2 \\
&=\; 1 - \tfrac{1}{3} n^{-1}\, t^2 \;\le\; \exp\{-\tfrac{1}{3} n^{-1}\, t^2\}.
\end{aligned} \tag{3.18}$$

Here $1 \ge \frac{1}{2} n^{-1}\, t^2$ because $t^2 \le a^2 \le \tilde{\beta}_3^{-2}\, n \le n$. Clearly we have the same bound for $|\mathbb{E}\, e\{tZ_j\}|$. Now for $j = 1, \ldots, n$ we define

$$U_j \;:=\; \textstyle\sum_{k=1}^{j-1} T_k + \sum_{k=j+1}^{n} Z_k. \tag{3.19}$$

We have that

$$\begin{aligned}
|\mathbb{E}\, e\{t\mathbb{T}_1\} - \mathbb{E}\, e\{tZ\}| \;&\le\; \textstyle\sum_{j=1}^n |\mathbb{E}\, e\{t(U_j + T_j)\} - \mathbb{E}\, e\{t(U_j + Z_j)\}| \\
&=\; \textstyle\sum_{j=1}^n |\mathbb{E}\, e\{tT_j\} - \mathbb{E}\, e\{tZ_j\}|\, |\mathbb{E}\, e\{tU_j\}|.
\end{aligned} \tag{3.20}$$

Now by (3.18) and the remark following it,

$$|\mathbb{E}\, e\{tU_j\}| \;\le\; \exp\{-\tfrac{1}{3}(1 - n^{-1})\, t^2\} \;\le\; \exp\{-\tfrac{1}{4} t^2\}.$$

Note that it is as well clear that

$$|\mathbb{E}\, e\{t(\mathbb{T}_1 - (T_1 + T_2))\}| \;\le\; \exp\{-\tfrac{1}{3}(1 - \tfrac{2}{n})\, t^2\} \;\le\; \exp\{-\tfrac{1}{4} t^2\}. \tag{3.21}$$

Using as well (3.16) and the remark following it we see that

$$\begin{aligned}
\delta_1 \;&\le\; \int_0^a \tilde{l}_n(t)\; \exp\{-\tfrac{1}{4} t^2\}\, \tfrac{1}{6} n\, t^3\, (\mathbb{E}\,|T_1|^3 + \mathbb{E}\,|Z_1|^3)\, t^{-1}\, dt \\
&\ll\; \tilde{\beta}_3\, n^{-1/2} \int_0^\infty t^2\, \tilde{l}_n(t)\; \exp\{-\tfrac{1}{4} t^2\}\, dt \;\ll\; \tilde{\beta}_3\, n^{-1/2} \;\le\; a^{-1},
\end{aligned}$$

which shows us that indeed $\hat{g} \sim \hat{h}_1$.

Comparing in turn $\hat{h}_1$ to $\hat{h}_2$, let $\delta_2 := \int_0^a |\hat{h}_1(t) - \hat{h}_2(t)|\, t^{-1}\, dt$. Then

$$\delta_2 \;\le\; \int_0^a l_n(t)\, |\mathbb{E}\, e\{t(\mathbb{T}_1 - (T_1 + T_2))\}|\, |\mathbb{E}\, e\{t(T_1 + T_2)\} - 1|\, t^{-1}\, dt,$$

with $|\mathbb{E}\, e\{t(T_1 + T_2)\} - 1| \le \frac{1}{2} t^2\, \mathbb{E}\, |T_1 + T_2|^2 = n^{-1} t^2$, so that

$$\delta_2 \;\le\; n^{-1} \int_0^a t\, l_n(t)\, \exp\{-\tfrac{1}{4} t^2\}\, dt \;\ll\; n^{-1} \;\ll\; a^{-2},$$

cf. (3.21), and indeed $\hat{h}_1 \sim \hat{h}_2$. This completes the proof. □

Now we take care of the interval $[0, a^{1/(2k-2)}]$. Moreover we show that outside of this interval, the Edgeworth expansion does not play a role.

**Lemma 3.5.** *We have that*

$$\int_0^{a^{1/(2k-2)}} |\hat{f}(t) - \hat{h}_2(t)|\, t^{-1}\, dt \;\ll\; a^{-1}, \qquad \int_{a^{1/(2k-2)}}^a |\hat{h}_2(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

**Proof of Lemma 3.5.** The second bound is easy to obtain. Using (3.18) we see that $|\mathbb{E}\, e\{t\mathbb{T}_1\}| \le \exp\{-\frac{1}{3} t^2\}$, and using as well (3.21) we notice that

$$\begin{aligned}
\int_{a^{1/(2k-2)}}^a |\hat{h}_2(t)|\, t^{-1}\, dt &\le \int_{a^{1/(2k-2)}}^a t^{-1}\, \tilde{l}_n(t)\, \exp\{-\tfrac{1}{4} t^2\}\, dt \\
&= \int_{a^{1/(2k-2)}}^a t^{-(2k-2)}\, t^{2k-3}\, \tilde{l}_n(t)\, \exp\{-\tfrac{1}{4} t^2\}\, dt \\
&\le a^{-1} \int_0^\infty t^{2k-3}\, \tilde{l}_n(t)\, \exp\{-\tfrac{1}{4} t^2\}\, dt \;\ll\; a^{-1}.
\end{aligned}$$

Now let

$$\delta_3 \;:=\; \int_0^{a^{1/(2k-2)}} |\hat{f}(t) - \hat{h}_2(t)|\, t^{-1}\, dt.$$

We have the following:

$$\begin{aligned}
\hat{f}(t) &= \mathbb{E}\, e\{t\mathbb{T}_1\} + (it)\, \mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbb{T}_2 + R \\
&= \mathbb{E}\, e\{t\mathbb{T}_1\} + (it) \tbinom{n}{2}\, \mathbb{E}\, e\{t\mathbb{T}_1\}\, T_{1,2} + R \qquad (3.22) \\
&= \mathbb{E}\, e\{t\mathbb{T}_1\} + (it) \tbinom{n}{2}\, (\mathbb{E}\, e\{tT_1\}\, e\{tT_2\}\, T_{1,2})\; \mathbb{E}\, e\{t\,(\mathbb{T}_1 - (T_1 + T_2))\} + R,
\end{aligned}$$

with $|R| \le \frac{1}{2} t^2\, \mathbb{E}\, \mathbb{T}_2^2 \le \frac{1}{4} t^2\, (\tilde{\gamma}_2\, n^{-1})$ (see (A.24)). Next note that, for all $y \in \mathbb{R}$,

$$|e\{y\} - 1 - iy| \;\le\; |y|^{3/2}.$$

Indeed, by Taylor expanding we know that $|e\{y\} - 1 - iy| \le \frac{1}{2}y^2$, which is $\le |y|^{3/2}$ in case $|y| \le 4$, while in the case that $|y| > 4$,

$$|e\{y\} - 1 - iy| \;\le\; 2 + |y| \;\le\; 2 \cdot 4^{-3/2}|y|^{3/2} + \tfrac{1}{2}|y|^{3/2} \;\le\; |y|^{3/2}.$$

As a consequence, for $j = 1, 2$,

$$e\{tT_j\} \;=\; 1 + itT_j + R_j \tag{3.23}$$

with $|R_j| \le |t|^{3/2}\,|T_j|^{3/2}$. Using this, together with (2.17) and the remarks that follow it, we see that

$$\mathbb{E}\,e\{tT_1\}\,e\{tT_2\}\,T_{1,2} \;=\; (it)^2\,\mathbb{E}\,T_1T_2T_{1,2} + \tilde{R} \;=\; (it)^2\,n^{-5/2}\,\kappa + \tilde{R}, \tag{3.24}$$

with

$$\tilde{R} \;:=\; (it)\,\mathbb{E}\,T_1R_2T_{1,2} + (it)\,\mathbb{E}\,T_2R_1T_{1,2} + \mathbb{E}\,R_1R_2T_{1,2}.$$

Here

$$\begin{aligned}
|\tilde{R}| \;&\le\; 2\,t\,\mathbb{E}\,|T_1R_2T_{1,2}| + \mathbb{E}\,|R_1R_2T_{1,2}| \\
&\le\; 2\,t^{5/2}\,\mathbb{E}\,|T_1||T_2|^{3/2}|T_{1,2}| + t^3\,\mathbb{E}\,|T_1T_2|^{3/2}|T_{1,2}| \\
&\le\; 2\,n^{-11/4}\,t^{5/2}\,\{\mathbb{E}\,|n^{1/2}\,T_1|^2\,|n^{1/2}\,T_2|^3\}^{1/2}\,\{\mathbb{E}\,|n^{3/2}\,T_{1,2}|^2\}^{1/2} \\
&\qquad + n^{-3}\,t^3\,\{\mathbb{E}\,|n^{1/2}\,T_1|^3\,|n^{1/2}\,T_2|^3\}^{1/2}\,\{\mathbb{E}\,|n^{3/2}\,T_{1,2}|^2\}^{1/2} \\
&=\; n^{-11/4}\,t^{5/2}\,\tilde{\beta}_3^{1/2}\,\tilde{\gamma}_2^{1/2}\,(2 + n^{-1/4}\,t^{1/2}\,\tilde{\beta}_3^{1/2}),
\end{aligned}$$

whereas

$$n^{-1/4}\,t^{1/2}\,\tilde{\beta}_3^{1/2} \;=\; t^{1/2}\,\{\tilde{\beta}_3\,n^{-1/2}\}^{1/2} \le (t\,a^{-1})^{1/2} \;\le\; 1$$

and moreover

$$n^{-3/4}\,\tilde{\beta}_3^{1/2}\,\tilde{\gamma}_2^{1/2} \;=\; (\tilde{\beta}_3\,n^{-1/2})^{1/2}\,(\tilde{\gamma}_2\,n^{-1})^{1/2} \;\le\; a^{-1/2}\,a^{-1/2} = a^{-1},$$

so that

$$|\tilde{R}| \;\le\; 3\,n^{-2}\,t^{5/2}\,a^{-1}. \tag{3.25}$$

As a result of (3.22) and (3.24) we have that

$$\begin{aligned}
\hat{f}(t) - \hat{h}_2(t) \;&=\; R + \Big((it)^3\,\left\{\tbinom{n}{2} - \tfrac{1}{2}n^2\right\}\,n^{-5/2}\,\kappa + (it)\,\tbinom{n}{2}\,\tilde{R}\Big) \\
&\qquad\qquad\qquad\qquad \mathbb{E}\,e\{t\,(\mathbb{T}_1 - (T_1 + T_2))\} \\
&=\; R + (it)\,\mathbb{E}\,e\{t\,(\mathbb{T}_1 - (T_1 + T_2))\}\,\Big(\tbinom{n}{2}\,\tilde{R} - \tfrac{1}{2}(it)^2\,n^{-3/2}\,\kappa\Big)
\end{aligned}$$

so that with (3.21) and (3.25) we see that

$$\begin{aligned} \delta_3 &\le \tfrac{1}{4}(\tilde{\gamma}_2\, n^{-1}) \int_0^{a^{1/(2k-2)}} t\, dt \\ &\qquad + \int_0^{a^{1/(2k-2)}} \left(\tfrac{1}{2} n^2\, |\tilde{R}| + \tfrac{1}{2} n^{-3/2}\, |\kappa|\, t^2\right) \exp\{-\tfrac{1}{4}t^2\}\, dt \\ &\ll (\tilde{\gamma}_2\, n^{-1})\, a^{1/(k-1)} + \int_0^{a^{1/(2k-2)}} \left(a^{-1}\, t^{5/2} + \tilde{\gamma}_2^{1/2}\, n^{-3/2}\, t^2\right) \exp\{-\tfrac{1}{4}t^2\}\, dt \\ &\ll (\tilde{\gamma}_2\, n^{-1})\, a^{1/(k-1)} + a^{-1} + (\tilde{\gamma}_2\, n^{-1})^{1/2}\, n^{-1}. \end{aligned}$$

Using (3.14) and (3.15) we now see that $\delta_3 \ll a^{-1}$. This concludes the proof. $\square$

Let, for all $x \in \mathbb{R}$, $\lceil x \rceil := \min\{y \in \mathbb{Z} \,:\, y \ge x\}$. For convenience of notation we set

$$p \;:=\; \lceil \log(k-1)/\log 2 \rceil. \tag{3.26}$$

Then

$$\begin{aligned} p &= \min\{l \in \mathbb{N} :\, l \ge \log(k-1)/\log 2\} \;=\; \min\{l \in \mathbb{N} :\, 2^l \ge k-1\} \\ &= \min\{l \in \mathbb{N} :\, 2^{-(l+1)} \;\le\; 1/(2k-2)\}, \end{aligned}$$

that is, $p$ is the first natural number $l$ for which $2^{-(l+1)} \le 1/(2k-2)$.

In order to conclude the proof we need the following lemma:

**Lemma 3.6.** *Let $l \in \{0, 1, 2, \ldots, p\}$. We have that*

$$\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\hat{f}(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

**Proof of Lemma 3.6.** We go about in the same way as in Lemma 2.4, taking $m$ of the form

$$m \;:=\; [n\, a^{-r}], \quad \text{with } r := 2^{-l} - 1/(2k-2).$$

Note that $0 \le r \le 1$. We split up $\mathbb{T}$ as before, use the notations $M_j$ and $\varphi_j$ and condition on the event that $\varphi := \varphi_1 \le \delta\, n^{-1}$ for some $\delta > 0$. Here (2.15) leads us to the fact that the cost of the truncation is bounded by

$$\begin{aligned} \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{P}(\varphi > \delta\, n^{-1})\, t^{-1}\, dt &\ll (\log a)\, (\tilde{\gamma}_2\, n^{-1}) \\ &\le (\log a)\, a^{-1-1/(k-1)} \;\ll\; a^{-1}. \end{aligned} \tag{3.27}$$

Using a Taylor expansion we see that

$$\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\}|\, t^{-1}\, dt \;\leq\; \varepsilon_1 + \varepsilon_2 + \varepsilon_3$$

with

$$\begin{aligned} \varepsilon_1 &:= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\, I\{\varphi\}\, e\{t(\mathbb{T} - \mathbb{T}_{21})\}|\, t^{-1}\, dt, \\ \varepsilon_2 &:= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\, I\{\varphi\}\, e\{t(\mathbb{T} - \mathbb{T}_{21})\}\, \mathbb{T}_{21}|\, dt, \\ \varepsilon_3 &:= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \tfrac{1}{2} t\, \mathbb{E}\, \mathbb{T}_{21}^2\, dt. \end{aligned} \tag{3.28}$$

Note that Lemma 2.5 is telling us that

$$\mathbb{E}\, \mathbb{T}_{21}^2 \;\leq\; (mn^{-1})^2\, (\tilde{\gamma}_2\, n^{-1}),$$

so that (3.14) leads us to

$$\varepsilon_3 \;\ll\; (mn^{-1})^2\, (\tilde{\gamma}_2\, n^{-1})\, a^{2^{-l+1}} \;\leq\; a^{1/(k-1)}\, (\tilde{\gamma}_2\, n^{-1}) \;\leq\; a^{-1},$$

which means that $\varepsilon_3$ is small enough.

As in (2.9) we have that $|\mathbb{E}\, e\{tT_1\}| \leq 1 - \frac{1}{3} t^2\, n^{-1}$. Using the notation $Y = (X_{m+1}, \ldots, X_n)$ as in (2.12), and taking $\delta$ small enough, now as in (2.18) we see that

$$\begin{aligned} |\mathbb{E}\, (e\{t(T_1 + \tilde{M}_1)\}\, |\, Y)| &\leq 1 - t^2\, n^{-1} (\tfrac{1}{3} - \delta^{1/2} - \tfrac{1}{2}\delta) \\ &\leq \exp\{-\delta_1\, t^2\, n^{-1}\} \end{aligned} \tag{3.29}$$

for some $\delta_1 > 0$, and consequently, using the substitution $y^2 := \delta_2\, t^2\, a^{-r}$,

$$\begin{aligned} \varepsilon_1 &\leq \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \exp\{-\delta_1\, t^2\, mn^{-1}\}\, t^{-1}\, dt \\ &\leq \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \exp\{-\delta_2\, t^2\, a^{-r}\}\, t^{-1}\, dt \\ &\leq a^{-2^{-(l+1)}}\, \delta_3^{-1/2}\, a^{r/2} \int_{(\delta_3\, a^{-r})^{1/2}\, a^{2^{-(l+1)}}}^{\infty} \exp\{-y^2\}\, dy. \end{aligned} \tag{3.30}$$

Here $-\frac{1}{2} r + 2^{-(l+1)} = \frac{1}{4}(k-1)^{-1} > 0$, and as in (2.20) it follows that $\varepsilon_1 \ll a^{-1}$.

We turn to $\varepsilon_2$. As in (2.19) we see that

$$\begin{aligned} \varepsilon_2 &\leq \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{E}\, I\{\varphi\}\, |\mathbb{E}\, (e\{t(\mathbb{T} - \mathbb{T}_{21})\}\, \mathbb{T}_{21}\, |\, Y)|\, dt \\ &\leq \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{E}\, |\mathbb{E}\, (e\{t(\mathbb{T}_{11} + \tilde{\mathbb{T}}_{22})\}\, \mathbb{T}_{21}\, |\, Y)|\, dt \\ &= \binom{m}{2} \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{E}\, |Y_{12}|\, |\mathbb{E}\, (e\{t(T_1 + \tilde{M}_1)\}\, |\, Y)|^{m-2}\, dt, \end{aligned}$$

taking

$$Y_{12} \;:=\; \mathbb{E}\big(e\{t(T_1+\tilde{M}_1)\}\, e\{t(T_2+\tilde{M}_2)\}\, T_{1,2}\,|\,Y\big).$$

Here we have the Taylor expansion

$$e\{t(T_j+\tilde{M}_j)\} \;=\; 1+R_j$$

with $R_j = R_j(X_j, Y)$ and $|R_j| \le t\,|T_j+\tilde{M}_j|$ $(j=1,2)$, which is showing us that

$$\begin{aligned} |Y_{12}| \;=\; |\mathbb{E}\,(R_1 R_2\, T_{1,2}\,|\,Y)| \;&\le\; t^2\,\mathbb{E}\,(|T_1+\tilde{M}_1|\,|T_2+\tilde{M}_2|\,|T_{1,2}|\,|\,Y) \\ &\le\; t^2\,\mathbb{E}\,(|T_1+\tilde{M}_1|^2\,|\,Y)\,\{\mathbb{E}T_{1,2}^2\}^{1/2}, \qquad (3.31) \end{aligned}$$

where we used Hölder's inequality (conditionally). Using that $\varphi_1 \le \delta\, n^{-1}$ and that $T_1$ and $\tilde{M}_1$ are conditionally uncorrelated, this leads us to:

$$|Y_{12}| \;\le\; t^2\,(n^{-1}+\delta\, n^{-1})\, n^{-3/2}\,\tilde{\gamma}^{1/2} \;\ll\; n^{-2}\,(\tilde{\gamma}_2\, n^{-1})^{1/2}\, t^2.$$

Together with (3.29) it follows that

$$\varepsilon_2 \;\ll\; (mn^{-1})^2\,(\tilde{\gamma}_2\, n^{-1})^{1/2} \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^2\, \exp\{-\delta_1\,(m-2)\,n^{-1}\,t^2\}\,dt. \qquad (3.32)$$

Since $r \le 1$ we have that $m \ge [n\,a^{-1}] \ge [a] \ge 100$, so that $(m-2)\,n^{-1} \ge \delta_2\, m\, n^{-1}$ for some $\delta_2 > 0$, and in turn

$$(m-2)\,n^{-1} \;\ge\; \delta_2\,(a^{-r}-n^{-1}) \;\ge\; \delta_3\, a^{-r}$$

for some $\delta_3 > 0$ (as $n^{-1} \le a^{-2}$). Using as well the substitution $y^2 := a^{-r}\,t^2$, (3.32) now leads us to the fact that

$$\begin{aligned} \varepsilon_2 \;&\ll\; a^{-2r}\, a^{-\frac{1}{2}-\frac{1}{2}(k-1)^{-1}} \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^2\, \exp\{-\delta_1\,\delta_3\, a^{-r}\, t^2\}\,dt \\ &\le\; a^{-2r-\frac{1}{2}-\frac{1}{2}(k-1)^{-1}}\, a^{\frac{3}{2}r} \int_{a^{2^{-(l+1)}-r/2}}^{\infty} y^2\, \exp\{-\delta_1\,\delta_3\, y^2\}\,dy, \end{aligned}$$

and as for $\varepsilon_1$ we see that $\varepsilon_2 \ll a^{-1}$. This concludes the proof. □

A direct result of Lemma 3.6 is that

$$\int_{a^{2^{-(p+1)}}}^{a} |\hat{f}(t)|\, t^{-1}\,dt \;=\; \sum_{l=0}^{p} \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\hat{f}(t)|\, t^{-1}\,dt \;\ll\; a^{-1}.$$

As $a^{2^{-(p+1)}} \le a^{1/(2k-2)}$ then as well $\int_{a^{1/(2k-2)}}^{a} |\hat{f}(t)|\, t^{-1}\,dt \ll a^{-1}$, and with Lemma 3.4 and 3.5 it is clear that

$$\int_0^a |\hat{f}(t)-\hat{g}(t)|\, t^{-1}\,dt \;\ll\; a^{-1}.$$

This finishes the proof of Theorem 3.1 in the case of $U$-statistics of order 2.

## 3.3 Proof of Theorem 3.1 for arbitrary statistics

We go about as for $U$-statistics of order 2, using moreover methods and terminology from Section 2.4.

As in Section 3.2 we may assume that $\mathbb{E}\,\mathbb{T} = 0$, $s^2 = 1$ and $\varepsilon = k^{-1}$ for some integer $k \geq 3$. We take

$$a \;:=\; \min\{\tilde{\beta}_3^{-1}\,n^{1/2},\ (\tilde{\gamma}_2^{-1}\,n)^{1-\varepsilon},\ (\tilde{\delta}_2^{-1}\,n)^{1/2}\} \;\geq 100$$

and apply (3.11). Note that by definition of $a$,

$$\tilde{\gamma}_2\,n^{-1} \;\leq\; a^{-1-1/(k-1)} \quad \text{and} \quad \tilde{\delta}_2\,n^{-1} \;\leq\; a^{-2}. \tag{3.33}$$

The remainder terms $R_1$ and $R_2$ are dealt with as before. Moreover, the result of Lemma 3.4 is still applicable, so that we precisely have to prove that

$$\hat{f} \;\sim\; \hat{h}_2,$$

with $\hat{f}$ as before denoting the characteristic function of $\mathbb{T}$.

Taking some fixed integer $1 \leq m \leq n$, we use the decomposition

$$\mathbb{T} \;=\; \mathbb{T}_{11} + \mathbb{T}_{12} + \mathbb{T}_{21} + \mathbb{T}_{22} + \mathbb{T}_{23} + \textstyle\sum_{j=1}^{m}(M_j + N_j) + \Lambda_1 + \Lambda_2,$$

with $\mathbb{T}_1 = \mathbb{T}_{11} + \mathbb{T}_{12}$ as in (2.10), $\mathbb{T}_2 = \mathbb{T}_{21} + \mathbb{T}_{22} + \mathbb{T}_{23}$ as in (2.11), $M_j$ and $N_j$ as in (2.21),

$$\Lambda_1 := \textstyle\sum_{A:\,|A\cap\{1,\ldots,m\}|=2,\,|A|\geq 3} T_A \quad \text{and} \quad \Lambda_2 := \textstyle\sum_{A:\,|A\cap\{1,\ldots,m\}|\geq 3} T_A.$$

As in (2.22), let $\Lambda := \mathbb{T}_{21} + \Lambda_1 + \Lambda_2$. We need the following moment bounds (cf. Lemma 2.5):

**Lemma 3.7.** *We have:*

$$\mathbb{E}\,\Lambda_1^2 \;\leq\; \tfrac{1}{2}(mn^{-1})^2\,(\tilde{\delta}_2\,n^{-1}), \quad \mathbb{E}\,\Lambda_2^2 \;\leq\; \tfrac{1}{6}(mn^{-1})^2\,(\tilde{\delta}_2\,n^{-1}),$$

*and*

$$\mathbb{E}\,N_1^2 \;\leq\; n^{-1}\,(\tilde{\delta}_2\,n^{-1}).$$

**Proof of Lemma 3.7.** We use (A.22) extensively. As a consequence of it,

$$\begin{aligned} \mathbb{E}\Lambda_1^2 &= \binom{m}{2} \textstyle\sum_{\emptyset\neq A\subset\{m+1,\dots,n\}} \mathbb{E}T_{1,2,A}^2 \;=\; \binom{m}{2} \textstyle\sum_{l=3}^{n-m+2} \binom{n-m}{l-2} \mathbb{E}T_{1,\dots,l}^2 \\ &\le \tfrac{1}{2}m^2 \textstyle\sum_{l=3}^{n} \binom{n-2}{l-2} \mathbb{E}T_{1,\dots,l}^2 \;=\; \tfrac{1}{2}(mn^{-1})^2\,(\tilde{\delta}_2\, n^{-1}). \end{aligned}$$

On the other hand we have that

$$\mathbb{E}\Lambda_2^2 \;=\; \textstyle\sum_{l=3}^{n} A_l\; \mathbb{E}T_{1,\dots,l}^2,$$

by $A_l$ denoting the number of sets $A$ such that $|A| = l$ and $|A\cap\{1,\dots,m\}| \geq 3$. Any such $A$ has, say, $k$ elements in common with $\{1,\dots,m\}$, and $l-k$ with $\{m+1,\dots,n\}$, which can only happen under the condition that $3 \leq k \leq m$ and $0 \leq l-k \leq n-m$. Hence

$$\begin{aligned} A_l &= \textstyle\sum_{k=\max\{3,\,l-n+m\}}^{\min\{m,l\}} \binom{m}{k}\binom{n-m}{l-k} \\ &= \textstyle\sum_{k=\max\{3,\,l-n+m\}}^{\min\{m,l\}} \frac{m(m-1)}{k(k-1)} \binom{m-2}{k-2}\binom{n-m}{l-k} \\ &\le \tfrac{1}{6}m^2 \textstyle\sum_{r=\max\{1,\,l-n+m-2\}}^{\min\{m-2,\,l-2\}} \binom{m-2}{r}\binom{n-m}{l-2-r}, \end{aligned}$$

using the substitution $r := k-2$. Corresponding to the fact that instead of choosing $l-2$ elements from a set of size $n-2$, one may from the start divide up the latter into two groups of size $m-2$ and $n-m$ respectively and choose from them, we have

$$\textstyle\sum_{r=\max\{0,\,l-n+m-2\}}^{\min\{m-2,\,l-2\}} \binom{m-2}{r}\binom{n-m}{l-2-r} \;=\; \binom{n-2}{l-2} \tag{3.34}$$

(see Alberink and Pestman (1998), p. 184). As a result $A_l \leq \frac{1}{6}m^2\binom{n-2}{l-2}$, and, again using (A.22), indeed

$$\mathbb{E}\Lambda_2^2 \;\le\; \tfrac{1}{6}(mn^{-1})^2\,(\tilde{\delta}_2\, n^{-1}).$$

As to $N_1$, using (2.24) we see that

$$\begin{aligned} \mathbb{E}N_1^2 &\le n \textstyle\sum_{\emptyset\neq A\subset\{m+2,\dots,n\}} \mathbb{E}T_{1,m+1,A}^2 \\ &\le n \textstyle\sum_{A:\,\{1,2\}\subset A,\,|A|\geq 3} \mathbb{E}T_A^2 \;=\; n^{-1}\,(\tilde{\delta}_2\, n^{-1}), \end{aligned}$$

cf. (A.21). This finishes the proof. □

First we look at the interval $[0, a^{1/(2k-2)}]$.

**Lemma 3.8.** *We have that*

$$\int_0^{a^{1/(2k-2)}} |\hat{f}(t) - \hat{h}_2(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

**Proof of Lemma 3.8.** From the proof of Lemma 3.5 it is already clear that

$$\int_0^{a^{1/(2k-2)}} |\mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\} - \hat{h}_2(t)|\, t^{-1}\, dt \ \ll\ a^{-1}.$$

Taking $m = n$, we have that $\mathbf{T}_3 = \Lambda_2$, and using the foregoing lemma and (3.33) it is as well clear that

$$\int_0^{a^{1/(2k-2)}} |\hat{f}(t) - \mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\}|\, t^{-1}\, dt \ \le\ \delta_1 + \delta_2$$

with

$$\delta_1 \ :=\ \int_0^{a^{1/(2k-2)}} |\mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\}\, \mathbf{T}_3|\, dt$$

and

$$\begin{aligned} \delta_2 \ &:=\ \tfrac{1}{2} \int_0^{a^{1/(2k-2)}} t\, \mathbb{E}\, \mathbf{T}_3^2\, dt \\ &\le\ \tfrac{1}{12}\, (\tilde{\delta}_2\, n^{-1})\, a^{1/(k-1)} \ =\ \tfrac{1}{12}\, a^{-2+1/(k-1)} \ \ll\ a^{-1}. \end{aligned}$$

We turn to $\delta_1$. Using a simple Taylor expansion we see that

$$\delta_1 \ \le\ \int_0^{a^{1/(2k-2)}} |\mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbf{T}_3|\, dt + \int_0^{a^{1/(2k-2)}} t\, \mathbb{E}\, |\mathbb{T}_2|\, |\mathbf{T}_3|\, dt \ =:\ \delta_3 + \delta_4.$$

Here, using Hoeffding's inequality,

$$\begin{aligned} \delta_4 \ &\le\ \tfrac{1}{2} a^{1/(k-1)}\, (\mathbb{E}\, \mathbb{T}_2^2)^{1/2}\, (\mathbb{E}\, \mathbf{T}_3^2)^{1/2} \\ &\le\ \tfrac{1}{2} a^{1/(k-1)}\, (a^{-1-1/(k-1)})^{1/2}\, (a^{-2})^{1/2} \ =\ \tfrac{1}{2} a^{-\frac{3}{2}+\frac{1}{2}\frac{1}{k-1}} \ \ll\ a^{-1}, \end{aligned} \tag{3.35}$$

so that we may concentrate on $\delta_3$. We use an argument which is due to Van Zwet (1984) (cf. p. 432, formula (3.12)). To this, for all $t \in \mathbb{R}$, let

$$\gamma(t) \ :=\ \mathbb{E}\, e\{tT_1\} \quad \text{and} \quad \theta(t) \ :=\ |\gamma(t)|^2. \tag{3.36}$$

Using symmetry we see that

$$\begin{aligned} |\mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbf{T}_3| \ &\le\ \textstyle\sum_{l=3}^n \binom{n}{l}\, |\mathbb{E}\, e\{t\mathbb{T}_1\}\, T_{1,\dots,l}| \\ &=\ \textstyle\sum_{l=3}^n \binom{n}{l}\, |\mathbb{E}\, \prod_{j=l+1}^n e\{tT_j\}\, \prod_{j=1}^l e\{tT_j\}\, T_{1,\dots,l}| \\ &=\ \textstyle\sum_{l=3}^n \binom{n}{l}\, |\gamma(t)|^{n-l}\, |\mathbb{E}\, \prod_{j=1}^l (e\{tT_j\} - \gamma(t))\, T_{1,\dots,l}|, \end{aligned}$$

using as well mutual independence and the fact that, for example,

$$\mathbb{E}\,\gamma(t)\,\prod_{j=2}^{l} e\{tT_j\}\,T_{1,\dots,l} \;=\; \gamma(t)\,\mathbb{E}\,\prod_{j=2}^{l} e\{tT_j\}\,\mathbb{E}\,(T_{1,\dots,l}\,|\,X_2,\dots,X_l) \;=\; 0.$$

Now notice that

$$\binom{n}{l}^2 \;=\; \tfrac{(l+2)(l+1)}{(n+2)(n+1)}\binom{n+2}{l+2}\tfrac{n(n-1)}{l(l-1)}\binom{n-2}{l-2} \;\le\; 4\binom{n+2}{l+2}\binom{n-2}{l-2} \tag{3.37}$$

for $3 \le l \le n$, and

$$\begin{aligned}\mathbb{E}\,|e\{tT_1\}-\gamma(t)|^2 &= \mathbb{E}\,(e\{tT_1\}-\gamma(t))(\overline{e\{tT_1\}}-\overline{\gamma(t)}) \qquad (3.38)\\ &= 1-\gamma(t)\,\overline{\gamma(t)}-\gamma(t)\,\overline{\gamma(t)}+\gamma(t)\,\overline{\gamma(t)} \;=\; 1-\theta(t).\end{aligned}$$

Using Hoeffding's inequality, the Cauchy-Schwarz inequality and (A.22), this leads us to the fact that

$$\begin{aligned}&|\mathbb{E}\,e\{t\mathbb{T}_1\}\,\mathbf{T}_3|\\ &\le \sum_{l=3}^{n}\binom{n}{l}\,\theta(t)^{\frac12(n-l)}\,(\mathbb{E}\,|\prod_{j=1}^{l}(e\{tT_j\}-\gamma(t))|^2)^{1/2}\,(\mathbb{E}\,T_{1,\dots,l}^2)^{1/2}\\ &\le \sum_{l=3}^{n}\binom{n}{l}\,\theta(t)^{\frac12(n-l)}\,(1-\theta(t))^{\frac12 l}\,(\mathbb{E}\,T_{1,\dots,l}^2)^{1/2} \qquad (3.39)\\ &\ll \Big(\sum_{l=3}^{n}\binom{n+2}{l+2}\,\theta(t)^{n-l}\,(1-\theta(t))^{l}\Big)^{1/2}\Big(\sum_{l=3}^{n}\binom{n-2}{l-2}\,\mathbb{E}\,T_{1,\dots,l}^2\Big)^{1/2}\\ &= n^{-1}\,(\tilde\delta_2\,n^{-1})^{1/2}\,(1-\theta(t))^{-1}\Big(\sum_{l=3}^{n}\binom{n+2}{l+2}\,\theta(t)^{n-l}\,(1-\theta(t))^{l+2}\Big)^{1/2}.\end{aligned}$$

With Newton's binomial formula

$$\begin{aligned}\sum_{l=3}^{n}\binom{n+2}{l+2}\,\theta(t)^{n-l}\,(1-\theta(t))^{l+2} &= \sum_{r=5}^{n+2}\binom{n+2}{r}\,\theta(t)^{n+2-r}\,(1-\theta(t))^{r}\\ &\le \sum_{r=0}^{n+2}\binom{n+2}{r}\,\theta(t)^{n+2-r}\,(1-\theta(t))^{r}\\ &= (\theta(t)+(1-\theta(t)))^{n+2} \;=\; 1, \qquad (3.40)\end{aligned}$$

so that in fact

$$|\mathbb{E}\,e\{t\mathbb{T}_1\}\,\mathbf{T}_3| \;\ll\; n^{-1}\,a^{-1}\,(1-\theta(t))^{-1}. \tag{3.41}$$

We turn to $(1-\theta(t))^{-1}$. We take an independent copy $\overline{X}_1$ of $X_1$ and write $\hat T_1 := T_1(\overline{X}_1)$. Now

$$\theta(t) \;=\; \gamma(t)\,\overline{\gamma(t)} \;=\; \mathbb{E}\,e\{tT_1\}\,\mathbb{E}\,e\{-t\hat T_1\} \;=\; \mathbb{E}\,e\{t(T_1-\hat T_1)\}. \tag{3.42}$$

A simple Taylor expansion leads us to:

$$\theta(t) \;=\; 1+(it)\,\mathbb{E}\,(T_1-\hat T_1)+\tfrac12(it)^2\,\mathbb{E}\,|T_1-\hat T_1|^2+R \;=\; 1-n^{-1}\,t^2+R,$$

with $|R| \le \frac{1}{6}t^3\,\mathbb{E}\,|T_1 - \hat{T}_1|^3 \le \frac{1}{6}a\,t^2\,\mathbb{E}\,|T_1-\hat{T}_1|^3$. With Lyapunov's inequality

$$\begin{aligned} \mathbb{E}\,|T_1 - \hat{T}_1|^3 &\le \mathbb{E}\,(T_1 - \hat{T}_1)^2\,(|T_1| + |\hat{T}_1|) \;=\; 2\,\mathbb{E}\,|T_1|\,(T_1^2 - 2T_1\hat{T}_1 + \hat{T}_1^2) \\ &= 2\,(\mathbb{E}\,|T_1|^3 + \mathbb{E}\,|T_1|\,\mathbb{E}\,\hat{T}_1^2) \;\le\; 4n^{-1}\,(\tilde{\beta}_3\,n^{-1/2}) \;\le\; 4n^{-1}\,a^{-1}, \end{aligned}$$

so that

$$1 - \theta(t) \;=\; n^{-1}\,t^2 - R \;\ge\; n^{-1}\,t^2\,(1 - \tfrac{1}{6}\cdot 4) \;=\; \tfrac{1}{3}n^{-1}\,t^2, \tag{3.43}$$

and from (3.41) it follows that

$$|\mathbb{E}\,e\{t\mathbb{T}_1\}\,\mathbf{T}_3| \;\ll\; a^{-1}\,t^{-2}.$$

As a consequence

$$\begin{aligned} \delta_3 &\ll \int_0^1 \mathbb{E}\,|\mathbf{T}_3|\,dt + a^{-1}\int_1^{a^{1/(2k-2)}} t^{-2}\,dt \\ &\le (\mathbb{E}\,\mathbf{T}_3^2)^{1/2} + a^{-1} \;\le\; 2a^{-1}. \end{aligned} \tag{3.44}$$

This finishes the proof. □

Lemma 3.5 is telling us that the Edgeworth expansion does not play any role outside of the interval $[0, a^{1/(2k-2)}]$, so that we may finally concentrate on finding a bound for

$$\int_{a^{1/(2k-2)}}^{a} |\hat{f}(t)|\,t^{-1}\,dt.$$

We proceed as in Lemma 3.6.

**Lemma 3.9.** *Let $p := \lceil \log(k-1)/\log 2 \rceil$. For any $l \in \{0,1,2,\ldots,p\}$, we have that*

$$\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\hat{f}(t)|\,t^{-1}\,dt \;\ll\; a^{-1}. \tag{3.45}$$

**Proof of Lemma 3.9.** We go about as in the proof of Lemma 3.6, but need to take care of the extra terms $\Lambda_1$, $\Lambda_2$ and $\sum_{j=1}^m N_j$. Let again

$$m \;:=\; [n\,a^{-r}] \quad \text{with } r := 2^{-l} - 1/(2k-2).$$

In order to deal with the $N_j$ we adjust the truncation. In fact we take

$$\varphi \;=\; \varphi_1 \;:=\; \mathbb{E}\,(M_1^2 + N_1^2\,|\,Y)$$

and use the indicator $I\{\varphi\} := I\{\varphi \leq \delta\, n^{-1}\}$. Using as well the result of Lemma 3.7, here we have that

$$\begin{aligned} \mathbb{E}\,\varphi &\leq n\,\mathbb{E}\,T_{1,2}^2 + n^{-1}\,(\tilde{\delta}_2\, n^{-1}) \\ &= n^{-1}\,(\tilde{\gamma}_2\, n^{-1}) + n^{-1}\,(\tilde{\delta}_2\, n^{-1}) \;\leq\; n^{-1}\, a^{-1-1/(k-1)}, \end{aligned}$$

and as in (3.27) the cost of the truncation is bounded by $a^{-1}$.

Next we derive bounds for the $\varepsilon_1$, $\varepsilon_2$ and $\varepsilon_3$ from (3.28), with $\mathbb{T}_{21}$ replaced by $\Lambda = \mathbb{T}_{21} + \Lambda_1 + \Lambda_2$. By the Lemmas 2.5 and 3.7 we have that

$$\mathbb{E}\,\Lambda^2 \;\ll\; (mn^{-1})^2\,(\tilde{\Delta}_2^2\, n^{-1}) \;\ll\; a^{-2r}\, a^{-1-1/(k-1)},$$

so that

$$\varepsilon_3 \;=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \tfrac{1}{2} t\,\mathbb{E}\,\Lambda^2\, dt \;\ll\; a^{-1}.$$

We take again $Y = (X_{m+1}, \dots, X_n)$. As to $\varepsilon_1$, letting $\tilde{M}_j := I\{\varphi\}\, M_j$ and $\tilde{N}_j := I\{\varphi\}\, N_j$, as in (3.29) we see that

$$\begin{aligned} |\mathbb{E}\,(e\{t(T_1 + \tilde{M}_1 + \tilde{N}_1)\}\,|\,Y)| &\leq 1 - t^2\, n^{-1}(\tfrac{1}{3} - 2\delta^{1/2} - \delta) \\ &\leq \exp\{-\delta_1\, t^2\, n^{-1}\}, \end{aligned}$$

for some $\delta_1 > 0$, which as in (3.30) leads us to the fact that

$$\varepsilon_1 \;=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\, I\{\varphi\}\, e\{t(\mathbb{T} - \Lambda)\}|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

As to $\varepsilon_2$ we have that $\varepsilon_2 \leq \varepsilon_{21} + \varepsilon_{22}$, with

$$\begin{aligned} \varepsilon_{21} &:= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\, I\{\varphi\}\, e\{t(\mathbb{T} - \Lambda)\}\, \mathbb{T}_{21}|\, dt, \\ \varepsilon_{22} &:= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\, I\{\varphi\}\, e\{t(\mathbb{T} - \Lambda)\}\, (\Lambda_1 + \Lambda_2)|\, dt. \end{aligned}$$

As in the proof of Lemma 3.6 we easily see that $\varepsilon_{21} \ll a^{-1}$. As to $\varepsilon_{22}$ we go about as in the proof of Lemma 3.8.

To this, let, for $1 \leq j \leq m$ and $3 \leq k \leq m$,

$$U_j \;:=\; T_j + \tilde{M}_j + \tilde{N}_j \quad \text{and} \quad V_k \;:=\; \sum\nolimits_{A \subset \{m+1, \dots, n\}} T_{1,\dots,k,A},$$

and, for all $t \in \mathbb{R}$,

$$\gamma(t) \;:=\; \mathbb{E}\,(e\{tU_1\}\,|\,Y) \quad \text{and} \quad \theta(t) \;:=\; |\gamma(t)|^2.$$

We condition on $Y = (X_{m+1}, \dots, X_n)$ as in Lemma 3.6, and use the approach of Lemma 3.8. Thus we see that

$$\begin{aligned}
\varepsilon_{22} &\le \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{E}\,|\mathbb{E}\,(e\{t\textstyle\sum_{j=1}^m U_j\}\ \sum_{A:\,|A\cap\{1,\dots,m\}|\ge 3} T_A \,|\, Y)|\,dt \\
&\le \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \textstyle\sum_{k=3}^m \binom{m}{k}\,\mathbb{E}\,|\mathbb{E}\,(e\{t\sum_{j=1}^m U_j\}\,V_k \,|\, Y)|\,dt \\
&= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \textstyle\sum_{k=3}^m \binom{m}{k}\,\mathbb{E}\,|\gamma(t)|^{m-k}\,|\mathbb{E}\,(\prod_{j=1}^k (e\{tU_j\} - \gamma(t))\,V_k \,|\, Y)|\,dt \\
&= \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \textstyle\sum_{k=3}^m \binom{m}{k}\,\mathbb{E}\,\theta(t)^{\frac{1}{2}(m-k)}\,W^{k/2}\,\mathbb{E}\,(V_k^2 \,|\, Y)^{1/2}\,dt,
\end{aligned}$$

denoting

$$W \;=\; W(Y) \;:=\; \mathbb{E}\,(|e\{tU_1\} - \gamma(t)|^2 \,|\, Y).$$

Here

$$W \;=\; 1 - \theta(t)$$

as in (3.38), and $\binom{m}{k} \le 2\binom{m+2}{k+2}^{1/2}\binom{m-2}{k-2}^{1/2}$ as in (3.37), so using Hölder's inequality and the Cauchy-Schwarz inequality we see that

$$\begin{aligned}
\varepsilon_{22} &\ll \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \textstyle\sum_{k=3}^m \binom{m+2}{k+2}^{1/2}\binom{m-2}{k-2}^{1/2}\left(\mathbb{E}\,\theta(t)^{m-k}\,(1-\theta(t))^k\right)^{\frac{1}{2}}(\mathbb{E}\,V_k^2)^{1/2}\,dt \\
&\le \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \left(\mathbb{E}\,\tilde{W}\right)^{1/2}\left(\textstyle\sum_{k=3}^m \binom{m-2}{k-2}\,\mathbb{E}\,V_k^2\right)^{1/2} dt, \qquad (3.46)
\end{aligned}$$

denoting

$$\tilde{W} \;=\; \tilde{W}(Y) \;:=\; \textstyle\sum_{k=3}^m \binom{m+2}{k+2}\,\theta(t)^{m-k}\,(1-\theta(t))^k.$$

As in (3.40) we have that

$$\tilde{W} \;\le\; (1-\theta(t))^{-2}, \qquad (3.47)$$

whereas moreover

$$\begin{aligned}
\textstyle\sum_{k=3}^m \binom{m-2}{k-2}\,\mathbb{E}\,V_k^2 &= \textstyle\sum_{k=3}^m \sum_{A\subset\{m+1,\dots,n\}} \binom{m-2}{k-2}\,\mathbb{E}\,T_{1,\dots,k,A}^2 \\
&= \textstyle\sum_{l=3}^n \sum_{k=\max\{3,\,l-n+m\}}^{\min\{m,l\}} \binom{m-2}{k-2}\binom{n-m}{l-k}\,\mathbb{E}\,T_{1,\dots,l}^2\,.
\end{aligned}$$

Using (A.22) and (3.34), from this we see that

$$\textstyle\sum_{k=3}^m \binom{m-2}{k-2}\,\mathbb{E}\,V_k^2 \;\le\; \sum_{l=3}^n \binom{n-2}{l-2}\,\mathbb{E}\,T_{1,\dots,l}^2 \;=\; n^{-2}\,(\tilde{\delta}_2\,n^{-1}) \;\le\; n^{-2}\,a^{-2}. \qquad (3.48)$$

As a result of (3.46), (3.47) and (3.48)

$$\varepsilon_{22} \ \ll \ n^{-1}\, a^{-1} \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} (\mathbb{E}\,(1-\theta(t))^{-2})^{1/2}\, dt, \tag{3.49}$$

and we finally concentrate on $\theta(t)$.

Let $\overline{X}_1$, as in Lemma 3.8, be an independent copy of $X_1$. Now let

$$\hat{U}_1 \ = \ \hat{T}_1 + \hat{M}_1 + \hat{N}_1 \ := \ T_1(\overline{X}_1) + \tilde{M}_1(\overline{X}_1) + \tilde{N}_1(\overline{X}_1).$$

As in (3.42) we have that

$$\theta(t) \ = \ \mathbb{E}\,(e\{t(U_1 - \hat{U}_1)\} \,|\, Y),$$

and using a simple Taylor expansion we see that

$$\theta(t) \ = \ \mathbb{E}\, e\{t(T_1 - \hat{T}_1)\} + R_1 + R_2, \tag{3.50}$$

with

$$R_1 \ := \ (it)\, \mathbb{E}\left(e\{t(T_1-\hat{T}_1)\}\,((\tilde{M}_1+\tilde{N}_1) - (\hat{M}_1+\hat{N}_1)) \,|\, Y\right)$$

and

$$\begin{aligned} |R_2| \ &\leq \ \tfrac{1}{2} t^2\, \mathbb{E}\left(|(\tilde{M}_1+\tilde{N}_1) - (\hat{M}_1+\hat{N}_1)|^2 \,|\, Y\right) \\ &= \ t^2\, \mathbb{E}\,(|\tilde{M}_1+\tilde{N}_1|^2 \,|\, Y) \ \leq \ \delta\, n^{-1}\, t^2. \end{aligned}$$

Now $\mathbb{E}\, e\{t(T_1-\hat{T}_1)\} \leq 1 - \frac{1}{3} n^{-1}\, t^2$, cf. (3.43), and moreover, using a Taylor expansion and Hölder's inequality,

$$\begin{aligned} |R_1| \ &\leq \ t^2\, \mathbb{E}\,(|T_1-\hat{T}_1|\, |(\tilde{M}_1+\tilde{N}_1) - (\hat{M}_1+\hat{N}_1)| \,|\, Y) \\ &\leq \ t^2\, (\mathbb{E}\,|T_1-\hat{T}_1|^2)^{1/2}\, \mathbb{E}\,(|(\tilde{M}_1+\tilde{N}_1) - (\hat{M}_1+\hat{N}_1)|^2 \,|\, Y)^{1/2} \\ &\leq \ t^2\, (2n^{-1})^{1/2}\, 2^{1/2}\, \mathbb{E}\,(|\tilde{M}_1+\tilde{N}_1|^2 \,|\, Y)^{1/2} \ \leq \ 2n^{-1}\, t^2\, \delta^{1/2}, \end{aligned}$$

so that in fact it follows from (3.50) that

$$1-\theta(t) \ = \ (1 - \mathbb{E}\, e\{t(T_1-\hat{T}_1)\}) - R_1 - R_2 \ \geq \ (\tfrac{1}{3} - 2\delta^{1/2} - \delta)\, n^{-1}\, t^2. \tag{3.51}$$

We take $\delta > 0$ such that $\delta_1 := \frac{1}{3} - 2\delta^{1/2} - \delta > 0$, after which $(1-\theta(t))^{-2} \leq (\delta_1^{-1}\, n\, t^{-2})^2$, and as a consequence of (3.49) and (3.51)

$$\varepsilon_{22} \ \ll \ n^{-1}\, a^{-1}\, \delta_1^{-1}\, n \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^{-2}\, dt \ \ll \ a^{-1}.$$

This finishes the proof. □

The result of Theorem 3.1 now easily follows from the Lemmas 3.4, 3.5, 3.8 and 3.9.

The substitution of $\sigma$ for $s$ in the theorem may be achieved by noting that (for example) either $s^2/\sigma^2 \geq \frac{9}{10}$, after which the new theorem is easily derived from the old one, or $s^2/\sigma^2 < \frac{9}{10}$, in which case, for example via (A.17), (A.25) and (A.26), $\frac{1}{2}(\tilde{\gamma}_2\, n^{-1} + \tilde{\delta}_2\, n^{-1}) \geq \frac{1}{10}\sigma^2$ and the theorem is trivially true (cf. (2.27) and what follows).

We conclude by giving the proof of Corollary 3.2.

**Proof of Corollary 3.2.** Let

$$\delta \;:=\; \sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\mathbb{T}} \leq x) - G(x)|, \tag{3.52}$$

and

$$b_1 := \tilde{\beta}_3\, s^{-3}\, n^{-1/2}, \quad b_2 := \tilde{\gamma}_2\, n^{-1}, \quad b_3 := (\tilde{\delta}_2\, n^{-1})^{1/2}. \tag{3.53}$$

Without loss of generality we assume that $s^2 = 1$ (again, later on we may just apply the reduced theorem to $\mathbb{T}/s$ and obtain the right result). Let $c$ denote the function that attaches to each pair $\varepsilon \in (0,1)$ 'the' implicit constant $c(\varepsilon)$ for which

$$\delta \;\leq\; c(\varepsilon)\, \max\{b_1, b_2^{1-\varepsilon}, b_3\}. \tag{3.54}$$

We choose $c \geq 1$ and decreasing in $\varepsilon$. This is sensible for the following reason. For any $\varepsilon^* \leq \varepsilon$ in $(0,1)$ we have that $\delta \leq c(\varepsilon^*)\, \max\{b_1, b_2^{1-\varepsilon^*}, b_3\}$. It follows that

$$\delta \;\leq\; c(\varepsilon^*)\, \max\{b_1,\, b_2^{1-\varepsilon},\, b_3\}. \tag{3.55}$$

in case $b_2 \leq 1$, while in the case that $b_2 > 1$ we have that

$$\delta \leq 1 \leq c(\varepsilon^*) \leq c(\varepsilon^*)\, \max\{b_1,\, b_2^{1-\varepsilon},\, b_3\},$$

and (3.55) is true as well. As a result we see that $c(\varepsilon^*)$ already functions as the implicit constant corresponding to $\varepsilon$, so it makes no sense to take $c(\varepsilon)$ larger than it.

Now let $b_4(\varepsilon) := \tilde\gamma_{2+\varepsilon}\, n^{-1}$. We show that

$$\delta \;\le\; c(\tilde\varepsilon)\,\max\{b_1,\, b_4(\varepsilon),\, b_3\}, \tag{3.56}$$

where $\tilde\varepsilon := \min\{\frac16, \frac14\varepsilon\}$. Note that $c(\tilde\varepsilon) \ge c(\frac16)$. As a consequence, if $b_2^{1-1/6} \le b_1$, that is to say, if $\tilde\gamma_2\, n^{-1} \le \tilde\beta_3^{6/5}\, n^{-3/5}$, with (3.54) we see that

$$\delta \;\le\; c(\tfrac16)\,\max\{b_1, b_3\} \;\le\; c(\tilde\varepsilon)\,\max\{b_1,\, b_3\}$$

and (3.56) is clearly true. Thus we may assume that $\tilde\gamma_2 \ge \tilde\beta_3^{6/5}\, n^{2/5} \ge n^{2/5}$. Now we have that $\tilde\gamma_{2+\varepsilon} \ge \tilde\gamma_2^{(2+\varepsilon)/2}$, whereas

$$\begin{aligned}\tilde\gamma_2^{(2+\varepsilon)/2} \ge n\, b_2^{1-\frac14\varepsilon} \;&\Leftrightarrow\; \tilde\gamma_2^{1+\frac12\varepsilon} \ge n\,\tilde\gamma_2^{1-\frac14\varepsilon}\, n^{-(1-\frac14\varepsilon)}\\ &\Leftrightarrow\; \tilde\gamma_2^{\frac34\varepsilon} \ge n^{\frac14\varepsilon} \;\Leftrightarrow\; \tilde\gamma_2 \ge n^{1/3},\end{aligned}$$

which is so by assumption. As a result $b_2^{1-\frac14\varepsilon} \le b_4(\varepsilon)$, and

$$\delta \;\le\; c(\tfrac14\varepsilon)\,\max\{b_1,\, b_4(\varepsilon),\, b_3\}$$

and again (3.56) is correct. This proves that we may replace $b_2^{1-\varepsilon}$ by $b_4(\varepsilon)$.

Now set $b_5(\varepsilon) := \tilde\gamma_2^{1+\varepsilon}\, n^{-1}$. We show that

$$\delta \;\le\; c(\hat\varepsilon)\,\max\{b_1,\, b_5(\varepsilon),\, b_3\}, \tag{3.57}$$

where $\hat\varepsilon := \min\{\frac16, \frac12\varepsilon\}$. As before, without loss of generality we may assume that $\tilde\gamma_2 \ge n^{2/5}$. But in that case $b_2^{1-\frac12\varepsilon} \le b_5(\varepsilon)$, the latter inequality being equivalent to $\tilde\gamma_2^{(1+\varepsilon)-(1-\frac12\varepsilon)} \ge n^{-(1-\frac12\varepsilon)+1}$, which is in turn equivalent to $\tilde\gamma_2 \ge n^{1/3}$, which is clearly true. As a result inequality (3.57) always holds, which shows that we may replace $b_2^{1-\varepsilon}$ by $b_5(\varepsilon)$.

This finishes the proof of the corollary. □

## 3.4 Proof of Theorem 3.3 for $U$-statistics of order 2

We turn to the proof of Theorem 3.3. First we look at the case of $U$-statistics of order 2. We follow the lines of the Sections 2.4 and 3.2.

As before we may assume that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$, and that $\varepsilon = k^{-1}$ for some natural number $k \geq 2$. We take

$$a \;:=\; \min\{\beta^{-1}, (\Delta_1^2)^{-(1-\varepsilon)}\} \;\geq\; 100$$

and are going to apply (3.12), with $G$ and $\hat{g}$ replaced by $H$ and $\hat{h}$. We write

$$k_n(t) \;:=\; (it)^3 \textstyle\sum_{1\leq j<k\leq n} \mathbb{E}\,T_j T_k T_{j,k} \quad \text{and} \quad \tilde{k}_n(t) \;:=\; 1 + k_n(t),$$

noting that

$$\hat{h}(t) \;=\; \tilde{k}_n(t)\,\exp\{-\tfrac{1}{2}t^2\}.$$

First we take a look at the remainder terms $R_1$ and $R_2$. Let as before $\hat{f}(t) = \mathbb{E}\,e\{t\mathbb{T}\}$. As to $R_1$, in Section 2.4 we have already proved that $\int_0^{\tilde{a}} |\hat{f}(t)|\,dt \ll 1$ for $\tilde{a} = \min\{\beta^{-1}, \Delta_1^{-2}\} \geq a$, so that

$$|R_1| \;\leq\; a^{-1} \textstyle\int_{-a}^{a} |\hat{f}(t)|\,dt \;\ll\; a^{-1}.$$

As to $R_2$, using (3.10) it is easily seen that

$$\begin{aligned} |R_2| \;&\leq\; a^{-1} \textstyle\int_{-a}^{a} |\tilde{k}_n(t)|\,\exp\{-\tfrac{1}{2}t^2\}\,dt \\ &\leq\; 2a^{-1} \textstyle\int_0^{a} (1 + t^3\,\Delta_1)\,\exp\{-\tfrac{1}{2}t^2\}\,dt \;\ll\; a^{-1}. \end{aligned} \tag{3.58}$$

This shows that our remainder terms are of the right order.

Now let $\hat{h}_1(t) := \tilde{k}_n(t)\,\mathbb{E}\,e\{t\mathbb{T}_1\}$ and

$$\hat{h}_2(t) \;:=\; \mathbb{E}\,e\{t\mathbb{T}_1\} + (it)^3 \sum_{1\leq j<k\leq n} (\mathbb{E}\,T_j T_k T_{j,k})\,\mathbb{E}\,e\{t(\mathbb{T}_1 - (T_j + T_k))\}.$$

We start by proving the following lemma, the analogue of Lemma 3.4, telling us that it makes no difference whether we use $\hat{h}$, $\hat{h}_1$ or $\hat{h}_2$ as our Edgeworth expansion.

**Lemma 3.10.** *We have that $\hat{h} \sim \hat{h}_1$ and that $\hat{h}_1 \sim \hat{h}_2$.*

**Proof of Lemma 3.10.** We go about as in the proof of Lemma 3.4. To this, let $Z_1, \ldots, Z_n$ be an independent sample, independent of the original sample, for which $Z_j$ is $N(0, s_j^2)$-distributed. Writing $Z := \sum_{j=1}^n Z_j$, $Z$ is again standard normally distributed. We set

$$\delta_1 \;:=\; \textstyle\int_0^a |\hat{h}(t) - \hat{h}_1(t)|\,t^{-1}\,dt$$

and prove that $\delta_1 \ll a^{-1}$. We take $U_j$ as in (3.19). Notice that

$$\begin{aligned} \delta_1 &= \int_0^a t^{-1}\,|\tilde{k}_n(t)|\,|\mathbb{E}\,e\{t\mathbb{T}_1\} - \mathbb{E}\,e\{tZ\}|\,dt \\ &\leq \int_0^a t^{-1}\,|\tilde{k}_n(t)|\,\sum_{j=1}^n |\mathbb{E}\,e\{tT_j\} - \mathbb{E}\,e\{tZ_j\}|\,|\mathbb{E}\,e\{tU_j\}|\,dt \end{aligned} \tag{3.59}$$

(see (3.20)). As to $\mathbb{E}\,e\{tT_j\}$ and $\mathbb{E}\,e\{tZ_j\}$ the Taylor expansions (3.16) and (3.17) apply, if we replace the factor $n^{-1}$ by $s_j^2$. This leads us to:

$$|\mathbb{E}\,e\{tT_j\} - \mathbb{E}\,e\{tZ_j\}| \;\leq\; \tfrac{1}{6}t^3\,(\mathbb{E}\,|T_j|^3 + \mathbb{E}\,|Z_j|^3) \;=\; \tfrac{1}{6}t^3\,(\beta_j + 4(2\pi)^{-1/2}\,s_j^3).$$

Notice that with Lyapunov's inequality $s_j^2 \leq \beta_j^{2/3}$, for $j = 1, \ldots, n$, so that

$$|\mathbb{E}\,e\{tT_j\} - \mathbb{E}\,e\{tZ_j\}| \;\leq\; \tfrac{1}{6}\,(1 + 4(2\pi)^{-1/2})\,\beta_j\,t^3, \tag{3.60}$$

whereas on the other hand necessarily $s_j^2 \leq \beta^{2/3} \leq a^{-2/3} \leq 100^{-2/3}$.

Moreover, (3.16) shows us as well that

$$|\mathbb{E}\,e\{tT_j\}| \;\leq\; |1 - \tfrac{1}{2}s_j^2\,t^2| + \tfrac{1}{6}t^3\,\beta_j.$$

In case $s_j^2 \leq 2\beta^2$ we have that $1 - \frac{1}{2}s_j^2\,t^2 \geq 0$ for all $t$, which leads to

$$|\mathbb{E}\,e\{tT_j\}| \;\leq\; 1 - \tfrac{1}{2}s_j^2\,t^2 + \tfrac{1}{6}t^2\,\beta_j\,\beta^{-1} \;\leq\; \exp\{-\tfrac{1}{2}t^2\,(s_j^2 - \tfrac{1}{3}\beta_j\,\beta^{-1})\},$$

whereas in the case that $s_j^2 > 2\beta^2$ we see that

$$s_j^2 \;\leq\; \beta_j^{2/3} \;\leq\; \beta_j^{2/3}\,2^{-\frac{1}{2}}\,s_j\,\beta^{-1} \;\leq\; 2^{-\frac{1}{2}}\beta_j\,\beta^{-1},$$

so that

$$|\mathbb{E}\,e\{tT_j\}| \;\leq\; 1 \;\leq\; \exp\{-\tfrac{1}{2}t^2\,(s_j^2 - 2^{-\frac{1}{2}}\,\beta_j\,\beta^{-1})\}.$$

As a result, in general we have the following bound:

$$|\mathbb{E}\,e\{tT_j\}| \;\leq\; \exp\{-\tfrac{1}{2}t^2\,(s_j^2 - 2^{-\frac{1}{2}}\,\beta_j\,\beta^{-1})\}. \tag{3.61}$$

Since $|\mathbb{E}\,e\{tZ_j\}| = \exp\{-\frac{1}{2}s_j^2\,t^2\}$ we have the same bound as in (3.61) for $Z_j$. As a result we see that

$$\begin{aligned} |\mathbb{E}\,e\{tU_j\}| &\leq \exp\{-\tfrac{1}{2}t^2\,\textstyle\sum_{k:k\neq j}(s_k^2 - 2^{-\frac{1}{2}}\,\beta_k\,\beta^{-1})\} \\ &\leq \exp\{-\tfrac{1}{2}t^2\,(1 - s_j^2 - 2^{-\frac{1}{2}})\} \end{aligned} \tag{3.62}$$

and via (3.59), (3.60) and (3.62), using as well (3.10), we see that

$$\begin{aligned} \delta_1 &\ll \int_0^a t^{-1}\,|\tilde{k}_n(t)|\,\sum_{j=1}^n t^3\,\beta_j\,\exp\{-\tfrac{1}{2}t^2\,(1-s_j^2-2^{-\frac{1}{2}})\}\,dt \\ &\ll \int_0^a t^{-1}\,(1+\Delta_1\,t^3)\,\beta\,t^3\,\exp\{-\tfrac{1}{2}t^2\,(1-100^{-2/3}-2^{-\frac{1}{2}})\}\,dt \\ &\ll \beta\int_0^\infty t^2\,\exp\{-\tfrac{1}{5}t^2\}\,(1+\Delta_1\,t^3)\,dt \;\ll\; \beta \;\le\; a^{-1}. \end{aligned}$$

This finishes the first part of the proof.

Comparing in turn $\hat{h}_1(t)$ to $\hat{h}_2(t)$, we show that

$$\delta_2 \;:=\; \int_0^a |\hat{h}_1(t)-\hat{h}_2(t)|\,t^{-1}\,dt \;\le\; a^{-1}.$$

To this, first we notice that

$$\begin{aligned} \delta_2 &\le \int_0^a \tfrac{1}{2}t^3\,\sum_{1\le j<k\le n}\mathbb{E}\,|T_jT_kT_{j,k}|\,|\mathbb{E}\,e\{t\mathbb{T}_1\}-V_{jk}|\,t^{-1}\,dt \\ &= \int_0^a \tfrac{1}{2}t^2\,\sum_{1\le j<k\le n}|V_{jk}|\,\mathbb{E}\,|T_jT_kT_{j,k}|\,|\mathbb{E}\,e\{t(T_j+T_k)\}-1|\,dt, \end{aligned}$$

denoting $V_{jk} := \mathbb{E}\,e\{t(\mathbb{T}_1-(T_j+T_k))\}$. Using (3.61) we see that

$$|V_{jk}| \;\le\; \exp\{-\tfrac{1}{2}t^2\,(1-(s_j^2+s_k^2)-2^{-\frac{1}{2}})\} \;\le\; \exp\{-\tfrac{1}{2}\,c_1\,t^2\}, \tag{3.63}$$

with $c_1 := 1-2^{-\frac{1}{2}}-2\cdot 100^{-2/3} > 0$. As by a short Taylor expansion

$$|\mathbb{E}\,e\{t(T_j+T_k)\}-1| \;\le\; \tfrac{1}{2}t^2\,\mathbb{E}\,|T_j+T_k|^2 \;=\; \tfrac{1}{2}t^2\,(s_j^2+s_k^2) \;\le\; t^2\,\beta^{2/3},$$

we get to the fact that (cf. (3.10))

$$\begin{aligned} \delta_2 &\le \tfrac{1}{2}\beta^{2/3}\int_0^a t^4\,\exp\{-\tfrac{1}{2}\,c_1\,t^2\}\sum_{1\le j<k\le n}\mathbb{E}\,|T_jT_kT_{j,k}|\,dt \\ &\le \tfrac{1}{2}\beta^{2/3}\,\Delta_1\int_0^\infty t^4\,\exp\{-\tfrac{1}{2}\,c_1\,t^2\}\,dt \\ &\ll \beta^{2/3}\,\Delta_1 \;\le\; \beta^{4/3}+\Delta_1^2 \;\ll\; a^{-1}. \end{aligned}$$

This finishes the second part of the proof. □

We turn to the analogue to Lemma 3.5, which tells us that the interval $[0, a^{1/(2k-2)}]$ is not important for our purposes and that outside of this interval the Edgeworth expansion is not important.

**Lemma 3.11.** *We have that*

$$\int_0^{a^{1/(2k-2)}} |\hat{f}(t)-\hat{h}_2(t)|\,t^{-1}\,dt \;\ll\; a^{-1}, \qquad \int_{a^{1/(2k-2)}}^a |\hat{h}_2(t)|\,t^{-1}\,dt \;\ll\; a^{-1}.$$

**Proof of Lemma 3.11.** Let $\delta_3 := \int_0^{a^{1/(2k-2)}} |\hat{f}(t) - \hat{h}_2(t)|\, t^{-1}\, dt$. As in (3.22),

$$\hat{f}(t) \;=\; \mathbb{E}\, e\{t\mathbb{T}_1\} + (it)\, \sum_{1\le j<k\le n} \mathbb{E}\, e\{t\mathbb{T}_1\}\, T_{j,k} + R,$$

with $|R| \le \frac{1}{2} t^2\, \mathbb{E}\, \mathbb{T}_2^2 \le \frac{1}{2} t^2\, a^{-1-1/(k-1)}$. As a result, see as well (3.63),

$$\begin{aligned} \delta_3 \;&\le\; \int_0^{a^{1/(2k-2)}} \Big| \sum_{1\le j<k\le n} \mathbb{E}\, e\{t(\mathbb{T}_1 - (T_j + T_k))\}\, W_{jk} \Big|\, dt \\ &\qquad\qquad + \int_0^{a^{1/(2k-2)}} t^{-1}\, |R|\, dt \\ &\le\; \tfrac{1}{4} a^{-1} + \int_0^{a^{1/(2k-2)}} \exp\{-\tfrac{1}{2}\, c_1\, t^2\} \sum_{1\le j<k\le n} |W_{jk}|\, dt, \end{aligned}$$

denoting

$$W_{jk} \;:=\; \mathbb{E}\, e\{t(T_j + T_k)\}\, T_{j,k} - (it)^2\, \mathbb{E}\, T_j T_k T_{j,k}.$$

Now we use the Taylor expansion $e\{tT_j\} \;=\; 1 + it\, T_j + R_j$ of (3.23), with $|R_j| \le |t|^{3/2}\, |T_j|^{3/2}$, which is showing us that

$$|W_{jk}| \;\le\; t\, \mathbb{E}\, |T_j R_k T_{j,k}| + t\, \mathbb{E}\, |R_j T_k T_{j,k}| + \mathbb{E}\, |R_j R_k T_{j,k}|.$$

Here

$$\mathbb{E}\, |T_j R_k T_{j,k}| \;\le\; t^{3/2}\, \{\mathbb{E}\, |T_j|^2 |T_k|^3\}^{1/2}\, \{\mathbb{E}\, T_{j,k}^2\}^{1/2} \;\le\; t^{3/2}\, \{s_j^2\, \beta_k\}^{1/2}\, \{\mathbb{E}\, T_{j,k}^2\}^{1/2}$$

and hence, with the Cauchy-Schwarz inequality,

$$\begin{aligned} \sum_{1\le j<k\le n} \mathbb{E}\, |T_j R_k T_{j,k}| \;&\le\; t^{3/2}\, \big(\sum_{1\le j<k\le n} s_j^2\, \beta_k\big)^{1/2}\, \big(\sum_{1\le j<k\le n} \mathbb{E}\, T_{j,k}^2\big)^{1/2} \\ &\le\; t^{3/2}\, \Delta_1\, \{(\sum_{j=1}^n s_j^2)\, (\sum_{k=1}^n \beta_k)\}^{1/2} \;=\; t^{3/2}\, \beta^{1/2}\, \Delta_1 \\ &\le\; t^{3/2}\, a^{-1/2}\, a^{-\frac{1}{2}(1+1/(k-1))} \;=\; t^{3/2}\, a^{-1-1/(2k-2)}. \end{aligned}$$

Of course in the same way $\sum_{1\le j<k\le n} \mathbb{E}\, |R_j T_k T_{j,k}| \;\le\; t^{3/2}\, a^{-1-1/(2k-2)}$, and finally

$$\begin{aligned} \sum_{1\le j<k\le n} \mathbb{E}\, |R_j R_k T_{j,k}| \;&\le\; t^3 \sum_{1\le j<k\le n} \{\mathbb{E}\, |T_j T_k|^3\}^{1/2}\, \{\mathbb{E}\, T_{j,k}^2\}^{1/2} \\ &\le\; t^3\, \big(\sum_{1\le j<k\le n} \beta_j\, \beta_k\big)^{1/2}\, (\Delta_1^2)^{1/2} \\ &\le\; t^3\, \big(\sum_{j=1}^n \beta_j\big)\, \Delta_1 \;\le\; t^3\, a^{-1-\frac{1}{2}(1+1/(k-1))}, \end{aligned}$$

and we see that

$$\delta_3 \le \tfrac{1}{4} a^{-1} + \int_0^{a^{1/(2k-2)}} \exp\{-\tfrac{1}{2}\, c_1\, t^2\}\, \big(2\, t^{5/2}\, a^{-1-1/(2k-2)} + t^3\, a^{-3/2-1/(2k-2)}\big)\, dt.$$

Clearly then $\delta_3 \leq a^{-1}$, which finishes the proof of the first statement.

The second bound is again easy to obtain. Using Lemma 3.10, we see that

$$\begin{aligned}
\int_{a^{1/(2k-2)}}^{a} |\hat{h}_2(t)|\, t^{-1}\, dt \;&\ll\; a^{-1} + \int_{a^{1/(2k-2)}}^{a} |\hat{h}(t)|\, t^{-1}\, dt \\
&\leq\; a^{-1} + \int_{a^{1/(2k-2)}}^{a} (1 + t^3\, \Delta_1)\, \exp\{-\tfrac{1}{2}t^2\}\, dt \\
&=\; a^{-1} + \int_{a^{1/(2k-2)}}^{a} t^{-(2k-2)}\, t^{2k-2}\, (1 + t^3\, \Delta_1)\, \exp\{-\tfrac{1}{2}t^2\}\, dt \\
&\leq\; a^{-1}\, (1 + \int_0^\infty t^{2k-2}\, (1 + t^3\, \Delta_1)\, \exp\{-\tfrac{1}{2}t^2\}\, dt) \;\ll\; a^{-1},
\end{aligned}$$

which proves the point. □

We still need to prove that

$$\int_{a^{1/(2k-2)}}^{a} |\hat{f}(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

We do this in two steps. First we take care of the interval $[a^{1/2}, a]$. After this we take care of the intervals $[a^{1/4}, a^{1/2}]$, $[a^{1/8}, a^{1/4}]$, …, until we reach $a^{1/(2k-2)}$.

**Lemma 3.12.** *We have that*

$$\int_{a^{1/2}}^{a} |\hat{f}(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

**Proof of Lemma 3.12.** We follow the same procedure of randomization as in the proof of Lemma 2.7. Instead of at $\mathbb{T}$ we look at $\tilde{\mathbb{T}} = \tilde{\mathbb{T}}_1 + \tilde{\mathbb{T}}_2$ as defined in (2.30), (2.31) and (2.32), that is, with

$$\tilde{\mathbb{T}} \;=\; \mathbb{T}\big(\alpha_1 X_1 + (1-\alpha_1)\overline{X}_1,\; \ldots\; , \alpha_n X_n + (1-\alpha_n)\overline{X}_n\big),$$

$\overline{X}_1, \ldots, \overline{X}_n$ constituting an independent copy of $X_1, \ldots, X_n$ and $\alpha_1, \ldots, \alpha_n$ an i.i.d. sample from a Bernoulli distributed population with

$$\mathbb{P}(\alpha_j = 1) \;=\; 1 - \mathbb{P}(\alpha_j = 0) \;=\; m,$$

independent of all other random variables. We now make the conceptual change of letting $m$ depend on $t$. To be exact, we take

$$m \;=\; m(t) \;:=\; \tilde{c}\, (\log t)\, t^{-2}$$

for some constant $\tilde{c} > 0$, thus for every $t$ obtaining a different statistic $\tilde{\mathbb{T}}$. Notice that still

$$\tilde{\mathbb{T}} \;=\; \tilde{\mathbb{T}}(t) \;\stackrel{d}{=}\; \mathbb{T},$$

for any fixed $t$, so that we may indeed look at $\tilde{\mathbb{T}}(t)$ instead of $\mathbb{T}$ everywhere. On the other hand, notice that $m(t) \in [0,1]$ for each $t \in [a^{1/2}, a]$ in case $\tilde{c} \le 2a \log^{-1} a$. In fact we take $\tilde{c} := 4\tilde{\delta}^{-1}$, with $\tilde{\delta}$ as in (2.43), whereas we shall see to it that $\tilde{\delta} \ge \frac{1}{10}$: then indeed $\tilde{c} \le 40 \le 2a \log^{-1} a$, and $m$ is defined correctly. We use the decomposition

$$\tilde{\mathbb{T}} \;=\; \tilde{\mathbb{T}}_1 + \tilde{\mathbb{T}}_2 \;=\; \tilde{\mathbb{T}}_{11} + \tilde{\mathbb{T}}_{12} + \tilde{\mathbb{T}}_{21} + \tilde{\mathbb{T}}_{22} + \tilde{\mathbb{T}}_{23}$$

as in (2.30).

Now we drop the tilde from our notation and look at

$$\delta_1 \;:=\; \int_{a^{1/2}}^{a} |\hat{f}(t)|\, t^{-1}\, dt.$$

Using a simple Taylor expansion, we see that $\delta_1 \le \delta_2 + \delta_3$ with

$$\delta_2 \;:=\; \int_{a^{1/2}}^{a} |\mathbb{E}\, e\{t(\mathbb{T} - \mathbb{T}_{21})\}|\, t^{-1}\, dt \quad \text{and} \quad \delta_3 \;:=\; \int_{a^{1/2}}^{a} \mathbb{E}\, |\mathbb{T}_{21}|\, dt. \tag{3.64}$$

Here

$$\begin{aligned} \delta_3 \;&\le\; \int_{a^{1/2}}^{a} \{\mathbb{E}\, \mathbb{T}_{21}^2\}^{1/2}\, dt \;=\; \int_{a^{1/2}}^{a} m(t)\, \Delta_1\, dt \;=\; \tilde{c}\, \Delta_1 \left[-t^{-1}\,(1 + \log t)\right]_{a^{1/2}}^{a} \\ &\le\; \tilde{c}\, a^{-\frac{1}{2}(1+1/(k-1))}\, a^{-\frac{1}{2}}\,(1 + \log a^{1/2}) \;\ll\; a^{-1}, \end{aligned}$$

so that we may further concentrate on $\delta_2$.

The estimation of $\delta_2$ goes as in the proof of Lemma 2.7. Let $M_j$, $\varphi_j$, $I\{\varphi\}$, $V_j$, $W_j$, $\gamma_j$, $\tilde{\alpha}_j$, $r_j$, $v_j$ and $w_j$ be as defined throughout the proof. Now, proceeding as in (2.34) and (2.36), we see that

$$\begin{aligned} \delta_2 \;&\le\; \int_{a^{1/2}}^{a} t^{-1}\, \mathbb{E} \prod_{j=1}^{n} V_j\, dt \\ &\le\; \int_{a^{1/2}}^{a} t^{-1}\, \mathbb{E}\, I\{\varphi\} \prod_{j=1}^{n} W_j\, dt + \int_{a^{1/2}}^{a} t^{-1}\, \mathbb{E}\,(1 - I\{\varphi\})\, dt \;=:\; \delta_4 + \delta_5. \end{aligned}$$

As in (2.37), $\mathbb{E}\,(1 - I\{\varphi\}) \le 2\delta^{-1}\, \Delta_1^2$ (for any fixed $t$), so

$$\delta_5 \;\le\; 2\delta^{-1}\, \Delta_1^2\, \log a \;\ll\; a^{-1}.$$

As to $\delta_4$ we use the precise same argumentation as in (2.38), (2.39) and what follows (for any fixed $t$), leading us to the fact that, cf. (2.42),

$$\delta_4 \;\le\; \int_{a^{1/2}}^{a} t^{-1}\, \exp\{-\tfrac{1}{2}\, \tilde{\delta}\, m(t)\, t^2\}\, dt,$$

so that

$$\delta_4 \;\le\; \int_{a^{1/2}}^{a} t^{-1-\tilde\delta\,\tilde c/2}\,dt \;=\; \big[-2\tilde\delta^{-1}\,\tilde c^{-1}\,t^{-\tilde\delta\,\tilde c/2}\big]_{a^{1/2}}^{a} \;\le\; 2\tilde\delta^{-1}\,\tilde c^{-1}\,a^{-\tilde\delta\,\tilde c/4}.$$

Taking $\tilde c := 4\tilde\delta^{-1}$ as proposed we see that $\delta_4 \le \frac{1}{2}a^{-1}$.

This concludes our proof in case $\tilde\delta \ge \frac{1}{10}$, but to this we just need to take $\delta > 0$ in such a way that

$$r_j < 0 \;\Rightarrow\; \gamma_j = 0, \quad \gamma_j = 0 \;\Rightarrow\; w_j \le 0 \quad \text{and} \quad \tilde\delta \ge \tfrac{1}{10},$$

which is possible. The proof is finished. □

Let $p$ as in (3.26) be the first natural number $r$ for which $2^{-(r+1)} \le 1/(2k-2)$.

**Lemma 3.13.** *Let $l \in \{1, 2, \ldots, p\}$. We have that*

$$\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\hat f(t)|\,t^{-1}\,dt \;\ll\; a^{-1}.$$

**Proof of Lemma 3.13.** The lemma is meaningless (and therefore true) in case $k = 2$ (which is so if and only if $p = 0$), so we assume that $k \ge 3$. Again we use the method of the proof of Lemma 2.7. The difference to our proceeding on the interval $[a^{1/2}, a]$ is that the short Taylor expansion that we were able to use in (3.64) no longer suffices for our purposes. Therefore we just take an expansion that is one term longer. The (conceptually) positive side of this is that we do not longer need to make $m$ dependent on $t$.

More concretely, let $\tilde{\mathbb{T}} = \tilde{\mathbb{T}}_1 + \tilde{\mathbb{T}}_2$ as before, with

$$m \;:=\; a^{-r} \quad \text{with} \quad r \;:=\; 2^{-l} - 1/(2k-2). \tag{3.65}$$

Note that $r \in [0, \frac{1}{2}]$. We prove that

$$\delta_1 \;:=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\,e\{t\tilde{\mathbb{T}}\}|\,t^{-1}\,dt \;\le\; a^{-1}.$$

We drop the tildes from the notation. Now $\delta_1 \le \delta_2 + \delta_3 + \delta_4$, taking

$$\begin{aligned}
\delta_2 \;&:=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\,e\{t(\mathbb{T} - \mathbb{T}_{21})\}|\,t^{-1}\,dt,\\
\delta_3 \;&:=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\,e\{t(\mathbb{T} - \mathbb{T}_{21})\}\,\mathbb{T}_{21}|\,dt
\end{aligned}$$

and

$$\begin{aligned}\delta_4 \;&:=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \tfrac{1}{2}\,(\mathbb{E}\,\mathbb{T}_{21}^2)\,t\,dt \;\le\; \tfrac{1}{2}m^2\,\Delta_1^2\,\tfrac{1}{2}a^{2^{-l+1}}\\ &\le\; \tfrac{1}{4}a^{-2^{l+1}+1/(k-1)}\,a^{-1-1/(k-1)}\,a^{2^{-l+1}} \;=\; \tfrac{1}{4}a^{-1}.\end{aligned}$$

We deal with $\delta_2$ as in the proofs of Lemma 2.7 and Lemma 3.12: this leads us to the bound

$$\begin{aligned}\delta_2 \;&\le\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^{-1}\,\exp\{-\tfrac{1}{2}\,\tilde\delta\,m\,t^2\}\,dt + \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^{-1}\,2\delta^{-1}\,\Delta_1^2\,dt\\ &\le\; \int_{a^{2^{-(l+1)}-r/2}}^{\infty} y^{-1}\,\exp\{-\tfrac{1}{2}\,\tilde\delta\,y^2\}\,dy + 2\delta^{-1}\,a^{-1-1/(k-1)}\,\log a^{2^{-l}},\end{aligned}$$

with $2^{-(l+1)} - \frac{1}{2}r = 1/(4k-4) > 0$, so that $\delta_2 \ll a^{-1}$. Note that while the implicit constant involved is dependent on $\delta$ and $\tilde\delta$, the latter two are not depending on the sample, and thus absolute constants.

Now we turn to $\delta_3$. Truncation leads to $\delta_3 \le \delta_5 + \delta_6$ with

$$\begin{aligned}\delta_5 \;&:=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\,I\{\varphi\}\,e\{t(\mathbb{T}-\mathbb{T}_{21})\}\,\mathbb{T}_{21}|\,dt,\\ \delta_6 \;&:=\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{E}\,(1-I\{\varphi\})\,|\mathbb{T}_{21}|\,dt,\end{aligned}$$

where, with the help of Hölder's inequality, and as in (2.37),

$$\begin{aligned}\delta_6 \;&\le\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \{\mathbb{E}\,(1-I\{\varphi\})\}^{1/2}\,\{\mathbb{E}\,\mathbb{T}_{21}^2\}^{1/2}\,dt\\ &\le\; (2\delta^{-1}\,\Delta_1^2)^{1/2}\,(m^2\,\Delta_1^2)^{1/2}\,a^{2^{-l}} \;\ll\; \delta^{-1/2}\,m\,\Delta_1^2\,a^{2^{-l}}\\ &\le\; \delta^{-1/2}\,a^{1/(2k-2)}\,a^{-1-1/(k-1)} \;\le\; \delta^{-1/2}\,a^{-1}.\end{aligned}$$

As to $\delta_5$,

$$\begin{aligned}\delta_5 \;&\le\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \textstyle\sum_{1\le j<k\le n} |\mathbb{E}\,I\{\varphi\}\,e\{t(\mathbb{T}-\mathbb{T}_{21})\}\,\alpha_j\alpha_k\,T_{j,k}|\,dt\\ &\le\; \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \textstyle\sum_{1\le j<k\le n} \mathbb{E}\,I\{\varphi\}\,\alpha_j\alpha_k\,|Y_{j,k}|\,|Z_{j,k}|\,dt, \qquad (3.66)\end{aligned}$$

with

$$\begin{aligned}Y_{j,k} \;&:=\; \mathbb{E}\,(e\{t(T_j+\tilde M_j)\}\,e\{t(T_k+\tilde M_k)\}\,T_{j,k}\,|\,\overline{X},\alpha),\\ Z_{j,k} \;&:=\; \textstyle\prod_{l\ne j,k} \mathbb{E}\,(e\{t\,\alpha_l\,(T_l+\tilde M_l)\}\,|\,\overline{X},\alpha).\end{aligned}$$

As to the second inequality in (3.66), note that

$$\begin{aligned}
&\mathbb{E}\big(I\{\varphi\}\, e\{t(\mathbb{T}-\mathbb{T}_{21})\}\,\alpha_j\alpha_k\, T_{j,k}\,|\,\overline{X},\alpha\big)\\
&\quad= e\{t(\mathbb{T}_{12}+\mathbb{T}_{23})\}\,\mathbb{E}\big(I\{\varphi\}\, e\{t\textstyle\sum_{l=1}^{n}\alpha_l\,(T_l+M_l)\}\,\alpha_j\alpha_k\, T_{j,k}\,|\,\overline{X},\alpha\big)\\
&\quad= e\{t(\mathbb{T}_{12}+\mathbb{T}_{23})\}\, I\{\varphi\}\,\alpha_j\alpha_k\,\mathbb{E}\big(e\{t(T_j+\tilde{M}_j)\}\, e\{t(T_k+\tilde{M}_k)\}\\
&\qquad\qquad T_{j,k}\, e\{t\textstyle\sum_{l\neq j,k}\alpha_l\,(T_l+\tilde{M}_l)\}\,|\,\overline{X},\alpha\big)\\
&\quad= e\{t(\mathbb{T}_{12}+\mathbb{T}_{23})\}\, I\{\varphi\}\,\alpha_j\alpha_k\, Y_{j,k}\, Z_{j,k}.
\end{aligned}$$

As in (3.31) we see that

$$\begin{aligned}
|Y_{j,k}| &\leq t^2\,\mathbb{E}\big(|T_j+\tilde{M}_j|\,|T_k+\tilde{M}_k|\,|T_{j,k}|\,|\,\overline{X},\alpha\big)\\
&\leq t^2\,\{s_j^2+\varphi_j^2\}^{1/2}\,\{s_k^2+\varphi_k^2\}^{1/2}\,\{\mathbb{E}T_{j,k}^2\}^{1/2}. \qquad (3.67)
\end{aligned}$$

As to $|Z_{j,k}|$ we have that

$$\begin{aligned}
I\{\varphi\}\,|Z_{j,k}| &\leq I\{\varphi\}\,\textstyle\prod_{l\neq j,k} W_l\\
&\leq \exp\{\tfrac{1}{2}\delta^{1/2}\,(1+\delta^{1/2})\,m\,t^2\}\,\exp\{-\tfrac{1}{2}t^2\textstyle\sum_{l\neq j,k}\tilde{\alpha}_l\, r_l\},
\end{aligned}$$

cf. (2.40). As $\tilde{\alpha}_j\, r_j+\tilde{\alpha}_k\, r_k\leq r_j+r_k\leq s_j^2+s_k^2\leq 2\beta^{2/3}$ this leads us to:

$$I\{\varphi\}\,|Z_{j,k}| \leq \exp\{\tfrac{1}{2}(\delta^{1/2}\,(1+\delta^{1/2})\,m+2\beta^{2/3})\,t^2\}\,\exp\{-\tfrac{1}{2}t^2\textstyle\sum_{l=1}^{n}\tilde{\alpha}_l\, r_l\}. \qquad (3.68)$$

Note that $\beta^{2/3}\leq a^{-2/3}=a^{-1/6}\,a^{-1/2}\leq 100^{-1/6}\,m$, which shows that the nuisance terms $\tilde{\alpha}_j\, r_j$ and $\tilde{\alpha}_k\, r_k$ do not bother us. Notice that this argument does not work for $l=0$, which is the precise reason why we had to take care of the interval $[a^{1/2},a]$ in a different way.

As a result of (3.66), (3.67) and (3.68),

$$\begin{aligned}
\delta_5 \leq &\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^2\,\exp\{\tfrac{1}{2}m\,t^2\,(\delta^{1/2}\,(1+\delta^{1/2})+2\cdot 100^{-1/6})\}\\
&\qquad \mathbb{E}\,\exp\{-\tfrac{1}{2}t^2\textstyle\sum_{l=1}^{n}\tilde{\alpha}_l\, r_l\}\, I\{\varphi\}\, U\, dt,
\end{aligned}$$

with

$$U := \textstyle\sum_{1\leq j<k\leq n}\{\alpha_j\,(s_j^2+\varphi_j^2)\}^{1/2}\,\{\alpha_k\,(s_k^2+\varphi_k^2)\}^{1/2}\,\{\mathbb{E}T_{j,k}^2\}^{1/2}.$$

Here the Cauchy-Schwarz inequality shows that

$$\begin{aligned}
I\{\varphi\}\, U &\leq I\{\varphi\}\,\big\{\textstyle\sum_{j<k}\alpha_j\,(s_j^2+\varphi_j^2)\,\alpha_k\,(s_k^2+\varphi_k^2)\big\}^{1/2}\,\big\{\textstyle\sum_{j<k}\mathbb{E}T_{j,k}^2\big\}^{1/2}\\
&\leq \Delta_1\, I\{\varphi\}\,\textstyle\sum_{j=1}^{n}\alpha_j\,(s_j^2+\varphi_j^2) \leq \Delta_1\,(1+\delta\, m),
\end{aligned}$$

(note that $\sum_j \alpha_j\, s_j^2 \le \sum_j s_j^2 = 1$), so that we may concentrate on the expression

$$\mathbb{E}\, \exp\{-\tfrac{1}{2}t^2 \sum_{l=1}^n \tilde{\alpha}_l\, r_l\},$$

which was already analyzed in (2.41). This leads us to:

$$\delta_5 \;\le\; \Delta_1\,(1+\delta\, m) \int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} t^2\, \exp\{-\tfrac{1}{2}\,\hat{\delta}\, m\, t^2\}\, dt$$

with $\hat{\delta} := \tilde{\delta} - 2\cdot 100^{-1/6}$ (taking $\tilde{\delta}$ as in (2.43)). Taking $\delta > 0$ such that

$$r_j < 0 \;\Rightarrow\; \gamma_j = 0, \qquad \gamma_j = 0 \;\Rightarrow\; w_j \le 0 \qquad \text{and} \qquad \tilde{\delta} > 0,$$

we now have that

$$\delta_5 \;\ll\; \Delta_1\, m^{-3/2} \int_{m^{1/2}\, a^{2^{-(l+1)}}}^{\infty} y^2\, \exp\{-\tfrac{1}{2}\,\hat{\delta}\, y^2\}\, dy,$$

and as $m^{1/2}\, a^{2^{-(l+1)}} = a^{1/(4k-4)}$, cf. (2.20),

$$\delta_5 \;\ll\; a^{-1}.$$

This finishes the proof. $\square$

Using the Lemmas 3.10, 3.11, 3.12 and 3.13, the proof of Theorem 3.3 confined to $U$-statistics of order 2 is easily concluded.

## 3.5 Proof of Theorem 3.3 for arbitrary statistics

We go about in the same way as for $U$-statistics of order 2. Again we may assume without loss of generality that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. First we assume that $\varepsilon = k^{-1}$ for some natural number $k \ge 2$. We take

$$a \;:=\; \min\{\beta^{-1}, (\Delta_1^2)^{-(1-\varepsilon)}, (\Delta_2^2)^{-\frac{1}{2}(1-\varepsilon)}\} \;\ge\; 100$$

and will apply (3.12), with $G$ and $\hat{g}$ replaced by $H$ and $\hat{h}$. Note that

$$\Delta_1^2 \;\le\; a^{-1-1/(k-1)}, \quad \Delta_2^2 \;\le\; a^{-2-2/(k-1)}.$$

We follow the lines of Section 3.4.

First we look at the remainder terms $R_1$ and $R_2$. In Section 2.5 it was already seen that $\int_0^{\tilde a} |\hat f(t)|\, dt \ll 1$ for some $\tilde a \geq a$: as a consequence,

$$|R_1| \;\ll\; a^{-1}.$$

Because of (3.58) we have that $|R_2| \ll a^{-1}$, and we may turn to proving that

$$\int_0^a |\hat f(t) - \hat h(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

Let $k_n$, $\tilde k_n$, $\hat h_1$, $\hat h_2$ be as in Section 3.4. As in Lemma 3.10 we see that $\hat h \sim \hat h_1$ and $\hat h_1 \sim \hat h_2$, so we may concentrate on $\hat h_2$ instead of $\hat h$.

We first look at the interval $[0, a^{1/(2k-2)}]$. Lemma 3.11 is telling us that

$$\int_0^{\tilde a^{1/(2k-2)}} |\mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\} - \hat h_2(t)|\, t^{-1}\, dt \;\ll\; \tilde a^{-1},$$

with $\tilde a \geq a$, so the result applies as well with $a$ instead of $\tilde a$. On the other hand,

$$\begin{aligned}\int_0^{a^{1/(2k-2)}} |\hat f(t) - \mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\}|\, t^{-1}\, dt \;&\leq\; \int_0^{a^{1/(2k-2)}} \mathbb{E}\, |\mathbf{T}_3|\, dt \\ &\leq\; a^{1/(2k-2)}\, (\Delta_2^2)^{1/2} \;\leq\; a^{-1},\end{aligned}$$

cf. (A.17), so that

$$\int_0^{a^{1/(2k-2)}} |\hat f(t) - \hat h_2(t)|\, t^{-1}\, dt \;\ll\; a^{-1},$$

and this solves our problem on the interval $[0, a^{1/(2k-2)}]$. As in the proof of Lemma 3.11 it is furthermore clear that

$$\int_{a^{1/(2k-2)}}^{a} |\hat h_2(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

We turn to the interval $[a^{1/2}, a]$. Here as in the proof of Lemma 3.12 we use a randomization procedure, where instead of at $\mathbb{T}$ we look at $\tilde{\mathbb{T}}$ as given in (2.44), taking

$$m \;=\; m(t) \;:=\; \tilde c\, (\log t)\, t^{-2},$$

with $\tilde c := 4\tilde\delta^{-1}$. We use the decomposition (2.44), moreover taking

$$\tilde\Lambda \;:=\; \tilde{\mathbb{T}}_{21} + \tilde\Lambda_1 + \tilde\Lambda_2,$$

with $\tilde{\mathbb{T}}_{21}$ as in (2.32),

$$\tilde{\Lambda}_1 \;:=\; \sum_{A:\,|A|=2} \alpha_A \sum_{B:\,A\cap B=\emptyset} (1-\alpha)_B\, T_{A,\overline{B}}$$

and

$$\tilde{\Lambda}_2 \;:=\; \sum_{A:\,|A|\geq 3} \alpha_A \sum_{B:\,A\cap B=\emptyset} (1-\alpha)_B\, T_{A,\overline{B}}.$$

Again we need bounds on the moments of $\tilde{\Lambda}_1$, $\tilde{\Lambda}_2$, the $\tilde{M}_j$ and the $\tilde{N}_j$. The bounds read as follows:

**Lemma 3.14.** *We have:*

$$\mathbb{E}\,\tilde{\Lambda}_1^2 \leq m^2\,\Delta_2^2, \quad \mathbb{E}\,\tilde{\Lambda}_2^2 \leq \tfrac{1}{3} m^2\,\Delta_2^2, \quad \sum_{j=1}^n \mathbb{E}\,\tilde{M}_j^2 \leq 2\,\Delta_1^2, \quad \sum_{j=1}^n \mathbb{E}\,\tilde{N}_j^2 \leq \Delta_2^2.$$

**Proof of Lemma 3.14.** Going about as in (A.18), we see that

$$\begin{aligned} \mathbb{E}\,\tilde{\Lambda}_1^2 \;&\leq\; m^2 \sum_{1\leq j<k\leq n} \sum_{B:\,j,k\notin B,|B|\geq 1} \mathbb{E}\,T^2_{j,k,\overline{B}} \\ &=\; m^2 \sum_{l=3}^n \binom{l}{2} \sum_{C:\,|C|=l} \mathbb{E}\,T_C^2 \;=\; m^2\,\Delta_2^2, \end{aligned}$$

see (A.19). On the other hand,

$$\begin{aligned} \mathbb{E}\,\tilde{\Lambda}_2^2 \;&=\; \sum_{A:\,|A|\geq 3} m^{|A|} \sum_{B:\,A\cap B=\emptyset} (1-m)^{|B|}\,\mathbb{E}\,T^2_{A,\overline{B}} \\ &=\; \sum_{l=3}^n \sum_{C:\,|C|=l} \Big\{ \sum_{k=3}^l \binom{l}{k} m^k\,(1-m)^{l-k} \Big\}\,\mathbb{E}\,T_C^2 \end{aligned} \tag{3.69}$$

(given any subset $C \subset N$ with $|C| = l$ (where $3 \leq l \leq n$), restricting ourselves to sets $A$ for which $|A| = k$ ($k = 3, \ldots, l$), we come $\binom{l}{k}$ times across $\mathbb{E}\,T_C^2$). We notice that

$$\begin{aligned} \sum_{k=3}^l \binom{l}{k} m^k\,(1-m)^{l-k} \;&=\; m^2 \sum_{k=3}^l \frac{l(l-1)}{k(k-1)} \binom{l-2}{k-2} m^{k-2}\,(1-m)^{l-k} \\ &\leq\; \tfrac{1}{3} m^2 \binom{l}{2} \sum_{r=1}^{l-2} \binom{l-2}{r} m^r\,(1-m)^{(l-2)-r} \;\leq\; \tfrac{1}{3} m^2 \binom{l}{2}, \end{aligned}$$

using the substitution $r := k-2$ and Newton's binomial formula, and from (3.69) and (A.19) it is clear that $\mathbb{E}\,\tilde{\Lambda}_2^2 \leq \frac{1}{3} m^2\,\Delta_2^2$.

As to the $\tilde{M}_j$'s, we see that

$$\sum_{j=1}^n \mathbb{E}\,\tilde{M}_j^2 \;\leq\; \sum_{j=1}^n \sum_{k:\,k\neq j} \mathbb{E}\,T^2_{j,\overline{k}} \;=\; 2\,\Delta_1^2.$$

Using the fact that $l \le \binom{l}{2}$ for $3 \le l \le n$, we finally see that

$$\begin{aligned} \textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{N}_j^2 &\le \textstyle\sum_{j=1}^n \sum_{B:\, j\notin B, |B|\ge 2} \mathbb{E}\,T_{j,\overline{B}}^2 \\ &= \textstyle\sum_{l=3}^n \sum_{C:\,|C|=l} l\,\mathbb{E}\,T_C^2 \;\le\; \Delta_2^2. \end{aligned}$$

This finishes the proof. □

We drop the tildes from our notation. As in (3.64) we remove the expression $\mathbb{T}_{21} + \Lambda_1 + \Lambda_2$ by means of a Taylor expansion, which is possible since

$$\mathbb{E}\,|\mathbb{T}_{21} + \Lambda_1 + \Lambda_2|^2 \;\le\; m^2\,(\Delta_1^2 + \tfrac{4}{3}\Delta_2^2) \;\ll\; m^2\,a^{-1-1/(k-1)}. \tag{3.70}$$

What remains is the term

$$\textstyle\int_{a^{1/2}}^a |\mathbb{E}\,e\{t(\mathbb{T}_{11} + \sum_{j=1}^n \alpha_j(M_j + N_j) + R)\}|\,t^{-1}\,dt,$$

which is estimated from above as in the proof of Lemma 2.7. We just need to take

$$I\{\varphi\} \;:=\; I\{\textstyle\sum_{j=1}^n \alpha_j\,\varphi_j^2 \le \delta\,m\},$$

with $\varphi_j^2 := \mathbb{E}\,(M_j^2 + N_j^2\,|\,\overline{X}, \alpha)$ as in (2.46), in order to proceed as before. The cost of introducing this truncation is bounded by

$$\begin{aligned} \textstyle\int_{a^{1/2}}^a (1 - \mathbb{E}\,I\{\varphi\})\,t^{-1}\,dt &\le \delta^{-1}\,m^{-1}\,m\,(2\Delta_1^2 + \Delta_2^2)\,\log a \\ &\ll a^{-1-1/(k-1)}\,\log a \;\ll\; a^{-1}, \end{aligned}$$

cf. (2.47), and thus small enough. The remaining part is dealt with in precisely the same way as in Lemma 3.12, and indeed

$$\textstyle\int_{a^{1/2}}^a |\hat{f}(t)|\,t^{-1}\,dt \;\ll\; a^{-1}.$$

Finally we turn to the analogue to Lemma 3.13. Again we switch from $\mathbb{T}$ to $\tilde{\mathbb{T}}$, with $m$ as in (3.65), and prove that

$$\textstyle\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} |\mathbb{E}\,e\{t\tilde{\mathbb{T}}\}|\,t^{-1}\,dt \;\ll\; a^{-1}.$$

We drop the tildes from our notation and remove the expression $\Lambda_1 + \Lambda_2$ at the cost of

$$\textstyle\int_{a^{2^{-(l+1)}}}^{a^{2^{-l}}} \mathbb{E}\,|\Lambda_1 + \Lambda_2|\,dt \;\ll\; (m^2\,\Delta_2^2)^{1/2}\,a^{2^{-l}} \;\ll\; a^{1/(2k-2)}\,a^{-1-1/(k-1)} \;\le\; a^{-1}.$$

Taking the indicator $I\{\varphi\}$ as on the interval $[a^{1/2}, a]$, the estimation is now concluded as in the proof of Lemma 3.13.

Normalization by $\sigma$ instead of $s$ is achieved as described at the end of the proof of Theorem 3.1 (for arbitrary statistics). This concludes the proof of Theorem 3.3. □

Finally we prove that, apart from the $\varepsilon$ in the power of $\Delta_2^2\, s^{-2}$, Theorem 3.3 is a generalization of Theorem 3.1. Indeed, in the case of an i.i.d. sample and a statistic that is symmetric in its arguments, from (2.3), (A.25) and (A.26) we already know that

$$\beta = \tilde{\beta}_3\, n^{-1/2}, \quad \Delta_1^2 \leq \tfrac{1}{2}\tilde{\gamma}_2\, n^{-1} \quad \text{and} \quad \Delta_2^2 \leq \tfrac{1}{2}\tilde{\delta}_2\, n^{-1}. \tag{3.71}$$

As always we may assume that

$$\tilde{\beta}_3\, s^{-3}\, n^{-1/2},\ \tilde{\gamma}_2\, s^{-2}\, n^{-1},\ \tilde{\delta}_2\, s^{-2}\, n^{-1} \ \leq\ 1, \tag{3.72}$$

since otherwise the trivial bound 1 will suffice. Then, for any $x \in \mathbb{R}$, using Hölder's inequality and (3.72) we see that

$$\begin{aligned} |G(x) - H(x)| &= |\tfrac{1}{2}n\, \Phi'''(x)\, s^{-3}\, \mathbb{E}T_1 T_2\, T_{1,2}| \\ &\leq \tfrac{1}{2}n\, |\Phi'''(x)|\, s^{-3}\, (\mathbb{E}T_1^2)\, \{\mathbb{E}T_{1,2}^2\}^{1/2} \\ &= \tfrac{1}{2}\, |\Phi'''(x)|\, n^{-3/2}\, (\tilde{\gamma}_2\, s^{-2})^{1/2} \ \leq\ \tfrac{1}{2}\, |\Phi'''(x)|\ n^{-1}. \end{aligned}$$

By Lyapunov's inequality $\tilde{\beta}_3 \geq s^3$, so moreover $n^{-1/2} \leq \tilde{\beta}_3\, s^{-3}\, n^{-1/2}$, and thus

$$\|G - H\|_\infty \ \leq\ \tfrac{1}{2}\, \|\Phi'''\|_\infty\, (\tilde{\beta}_3\, s^{-3}\, n^{-1/2})^2 \ \ll\ \big(\tilde{\beta}_3\, s^{-3}\, n^{-1/2}\big)^2. \tag{3.73}$$

Using (3.71) and (3.73), it is clear that then, in case the term $\varepsilon$ were not in the power of $\Delta_2^2\, s^{-2}$, Theorem 3.1 would indeed follow from Theorem 3.3.

# Chapter 4

# Applications

In the final chapter we look at what happens when we apply our main results, Theorems 2.2 and 3.3, to concrete examples. To be more exact, we will look in turn at simple linear rank statistics, general $U$-statistics, incomplete $U$-statistics, self-normalized statistics, Student's statistic and some self-normalized statistics associated with linear regression, for each of these finding concentration and Berry-Esseen bounds. The results are usually the classical ones, or a bit worse (in the case of Berry-Esseen bounds for self-normalized statistics). They are important not so much as to their content, but as illustrations of how our general method performs in practice. The fact that we are for example not able to obtain the classical Berry-Esseen bound for Student's statistic has in this respect significance as well.

Concentration bounds being not of that much importance in practice, not much effort has been put into obtaining them. To compare our concentration bounds to classical results, we use the following link to Berry-Esseen bounds. Let $\mathbb{T}$ be the statistic we are looking at. As in (3.1), in case $0 < s^2 < \infty$ we use the notation

$$D \;=\; D(\mathbb{T}) \;=\; \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left( \frac{\mathbb{T} - \mathbb{E}\,\mathbb{T}}{s} \le x \right) - \Phi(x) \right| . \tag{4.1}$$

Now suppose that it is known that, for some constant $c$ and moment expression $A$, we have that $D \le c\,A$. Taking $\tilde{\mathbb{T}} = (\mathbb{T} - \mathbb{E}\,\mathbb{T})/s$ as earlier, then

$$\begin{aligned} \mathbb{P}(x \le \tilde{\mathbb{T}} \le x+\lambda) \;&\le\; |\mathbb{P}(\tilde{\mathbb{T}} \le x+\lambda) - \Phi(x+\lambda)| + (\Phi(x+\lambda) - \Phi(x)) \\ &\qquad + |\Phi(x) - \mathbb{P}(\tilde{\mathbb{T}} < x)| \;\le\; 2c\,A + (2\pi)^{-1/2}\,\lambda, \end{aligned}$$

and thus

$$Q(\mathbb{T}/s, \lambda) \;=\; Q(\tilde{\mathbb{T}}, \lambda) \;\leq\; 2c\,A + (2\pi)^{-1/2}\,\lambda \;\ll\; \max\{\lambda, A\}.$$

In short, we have the following general implication:

$$D \;\ll\; A \quad\Rightarrow\quad Q(\mathbb{T}/s, \lambda) \;\ll\; \max\{\lambda, A\}, \tag{4.2}$$

which usually gives us a good idea about what we may expect as a concentration bound.

## 4.1 Simple linear rank statistics

Our first application will be on finding a Berry-Esseen bound for simple linear rank statistics, see e.g. Does (1982). After this we apply the result to Wilcoxon's rank-sum statistic.

Let $X_1, \ldots, X_n$ be independent, not necessarily identically distributed random variables, with distribution functions $F_1, \ldots, F_n$. We are going to test the null hypothesis that they are equal. We do this as follows. For $j = 1, \ldots, n$, we define the rank number of $X_j$ among $X_1, \ldots, X_n$ by

$$R_j \;:=\; \textstyle\sum_{i=1}^n I\{X_j \geq X_i\}.$$

We assume that the distribution functions are continuous, so that ties (equal outcomes) occur with probability 0. Taking certain real-valued coefficients $a_1, \ldots, a_n$, we look at the so-called simple linear rank statistic

$$\mathbb{T} \;:=\; \textstyle\sum_{j=1}^n a_j R_j. \tag{4.3}$$

Notice that this statistic is typically *not* symmetric in its arguments. Let $\overline{a} := n^{-1} \sum_{j=1}^n a_j$ and, for $p = 2, 3$,

$$\rho_p \;:=\; n^{-1} \textstyle\sum_{j=1}^n |a_j - \overline{a}|^p,$$

for which with Lyapunov's inequality $\rho_2^{3/2} \;\leq\; \rho_3$. We obtain the following result, which may be found in Does (1982):

**Theorem 4.1.** *Let $\mathbb{T}$ be a simple linear rank statistic, as defined in (4.3). Under the null hypothesis that $F_1 = \ldots = F_n$ and that the $F_j$ are continuous,* *we have that* $\mathbb{E}\,\mathbb{T} = \frac{1}{2}n(n+1)\,\overline{a}$, $s^2 = \frac{1}{12}n^3\,\rho_2$, *and an absolute constant* $c$ *exists such that*

$$D \;\leq\; c\;\rho_3\,\rho_2^{-3/2}\,n^{-1/2}.$$

**Proof of Theorem 4.1.** Under the null hypothesis $X_1, \ldots, X_n$ constitutes an i.i.d. sample. Let $X$ denote an independent copy of $X_1$. We are going to apply Theorem 3.3.

We may write

$$\mathbb{T} \;=\; \sum_{k=1}^n \sum_{l=1}^n a_k\, I\{X_k \geq X_l\} \;=\; n\,\overline{a} + \sum_{k=1}^n \sum_{l\neq k} a_k\, I\{X_k \geq X_l\}.$$

It is clear that $\mathbb{T}$ is a $U$-statistic of order 2, so that certainly $\Delta^2 = \Delta_1^2$, see (A.16). Now, for $j = 1, \ldots, n$, let

$$\xi_j \;:=\; \mathbb{E}\,(I\{X_j \geq X\}\,|\,X_j) - \tfrac{1}{2}.$$

Since with probability 1 there are no ties, we have that

$$\mathbb{E}\,(I\{X \geq X_j\}\,|\,X_j) - \tfrac{1}{2} \;=\; \big(1 - \mathbb{E}\,(I\{X_j \geq X\}\,|\,X_j)\big) - \tfrac{1}{2} \;=\; -\xi_j.$$

Writing $U := F(X)$ and $U_j := F(X_j)$, the random variables $U, U_1, \ldots, U_n$ are independent and uniformly distributed on $(0,1)$, so that

$$\mathbb{E}\xi_j^2 \;=\; \mathbb{E}\xi_1^2 \;=\; \mathbb{E}\,|\mathbb{E}\,(I\{U_1 \geq U\}\,|\,U_1) - \tfrac{1}{2}|^2 \;=\; \mathbb{E}\,|U_1 - \tfrac{1}{2}|^2 \;=\; \tfrac{1}{12},$$

and $\mathbb{E}\,|\xi_j|^3 = \mathbb{E}\,|\xi_1|^3 = \mathbb{E}\,|U_1 - \frac{1}{2}|^3 = \int_0^1 |y - \frac{1}{2}|^3\,dy = \frac{1}{32}$, for all $j$. Of course $\mathbb{E}\,\xi_j = 0$. Now first note that

$$\mathbb{E}\,\mathbb{T} \;=\; n\,\overline{a} + \sum_{k=1}^n \sum_{l\neq k} \tfrac{1}{2}a_k \;=\; n\,\overline{a} + \tfrac{1}{2}(n-1)\,n\,\overline{a} \;=\; \tfrac{1}{2}n(n+1)\,\overline{a}. \tag{4.4}$$

As to $T_j$, for $j = 1, \ldots, n$, we have the following:

$$\begin{aligned} T_j \;=\; \mathbb{E}\,(\mathbb{T}\,|\,X_j) - \mathbb{E}\,\mathbb{T} \;&=\; \sum_{k=1}^n \sum_{l\neq k} a_k\,\big(\mathbb{E}\,(I\{X_k \geq X_l\}\,|\,X_j) - \tfrac{1}{2}\big) \\ &=\; a_j \sum_{l\neq j} \xi_j + \sum_{k\neq j} a_k \cdot -\xi_j \qquad (4.5) \\ &=\; n\,a_j\,\xi_j - a_j\,\xi_j - \xi_j \sum_{k\neq j} a_k \;=\; n\,\xi_j\,(a_j - \overline{a}). \end{aligned}$$

Thus

$$s^2 \;=\; \sum_{j=1}^n \mathbb{E}\,T_j^2 \;=\; \sum_{j=1}^n n^2\,(a_j - \overline{a})^2\,\mathbb{E}\,\xi_j^2 \;=\; \tfrac{1}{12}n^3\,\rho_2, \tag{4.6}$$

and

$$\beta \;=\; \textstyle\sum_{j=1}^n \mathbb{E}\,|T_j|^3 \;=\; \sum_{j=1}^n n^3\,|a_j-\overline{a}|^3\,\mathbb{E}\,|\xi_j|^3 \;=\; \frac{1}{32}n^4\,\rho_3. \tag{4.7}$$

Furthermore, for all $1 \le i < j \le n$, we have that

$$\begin{aligned}\mathbb{E}\,(\mathbb{T}\,|\,X_i,X_j) - \mathbb{E}\,\mathbb{T} \;&=\; \textstyle\sum_{k=1}^n \sum_{l\neq k} a_k\,\big(\mathbb{E}\,(I\{X_k \ge X_l\}\,|\,X_i,X_j) - \frac12\big)\\ =\;& a_i\,(I\{X_i \ge X_j\} - \tfrac12) + a_j\,(I\{X_j \ge X_i\} - \tfrac12)\\ &+\; \textstyle\sum_{l\neq i,j} a_i\,\xi_i + \sum_{l\neq i,j} a_j\,\xi_j + \sum_{k\neq i,j} a_k\cdot -\xi_i + \sum_{k\neq i,j} a_k\cdot -\xi_j\\ =\;& (a_i-a_j)\,(I\{X_i \ge X_j\} - \tfrac12) + \xi_i\,((n-2)a_i - \textstyle\sum_{k\neq i,j} a_k)\\ &+\, \xi_j\,((n-2)a_j - \textstyle\sum_{k\neq i,j} a_k).\end{aligned}$$

Together with (4.5), this shows us that

$$\begin{aligned}T_{i,j} \;&=\; \big(\mathbb{E}\,(\mathbb{T}\,|\,X_i,X_j) - \mathbb{E}\,\mathbb{T}\big) - \big(\mathbb{E}\,(\mathbb{T}\,|\,X_i) - \mathbb{E}\,\mathbb{T}\big) - \big(\mathbb{E}\,(\mathbb{T}\,|\,X_j) - \mathbb{E}\,\mathbb{T}\big)\\ &=\; (a_i-a_j)\,(I\{X_i \ge X_j\} - \tfrac12)\\ &\qquad +\, \xi_i\,(-2a_i + a_i + a_j) + \xi_j\,(-2a_j + a_i + a_j)\\ &=\; (a_i-a_j)\,(I\{X_i \ge X_j\} - \xi_i + \xi_j - \tfrac12),\end{aligned}$$

and thus

$$\begin{aligned}\Delta_1^2 \;&=\; \textstyle\sum_{1\le i<j\le n}(a_i-a_j)^2\,\mathbb{E}\,|I\{X_i \ge X_j\} - \xi_i + \xi_j - \frac12|^2\\ &=\; \mathbb{E}\,|I\{X_1 \ge X_2\} - \xi_1 + \xi_2 - \tfrac12|^2\,\textstyle\sum_{1\le i<j\le n}(a_i-a_j)^2.\end{aligned}$$

Here

$$\begin{aligned}\textstyle\sum_{1\le i<j\le n}(a_i-a_j)^2 \;&=\; \tfrac12\textstyle\sum_{i=1}^n\sum_{j=1}^n(a_i-a_j)^2\\ &=\; \tfrac12\textstyle\sum_{i=1}^n\sum_{j=1}^n((a_i-\overline{a})-(a_j-\overline{a}))^2\\ &=\; 2\cdot\tfrac12\textstyle\sum_{i=1}^n(a_i-\overline{a})^2 + 2\cdot\frac12\sum_{i=1}^n\sum_{j=1}^n(a_i-\overline{a})(a_j-\overline{a})\\ &=\; n^2\,\rho_2 - \big(\textstyle\sum_{i=1}^n(a_i-\overline{a})\big)^2 \;=\; n^2\,\rho_2,\end{aligned}$$

whereas, taking $b := \mathbb{E}\,|I\{X_1 \ge X_2\} - \xi_1 + \xi_2 - \frac12|^2$,

$$\begin{aligned}b \;=\;& \mathbb{E}\,I\{X_1 \ge X_2\} - 2\,\mathbb{E}\,I\{X_1 \ge X_2\}\,\xi_1 + 2\,\mathbb{E}\,I\{X_1 \ge X_2\}\,\xi_2\\ &-\, 2\cdot\tfrac12\mathbb{E}\,I\{X_1 \ge X_2\} +\; \mathbb{E}\,\xi_1^2 - 2\,\mathbb{E}\,\xi_1\xi_2\\ &+\, 2\cdot\tfrac12\mathbb{E}\,\xi_1 + \mathbb{E}\,\xi_2^2 - 2\cdot\tfrac12\mathbb{E}\,\xi_2 + \tfrac14.\end{aligned}$$

Here as before

$$\begin{aligned}\mathbb{E}\, I\{X_1 \geq X_2\}\, \xi_1 \;&=\; \mathbb{E}\mathbb{E}\,(\xi_1\, I\{U_1 \geq U_2\}\,|\, U_1) \;=\; \mathbb{E}\xi_1\, U_1 \\ =\;\; \mathbb{E}U_1\,(\mathbb{E}\,(I\{U_1 \geq U\}\,|\,U_1) - \tfrac{1}{2}) \;&=\; \mathbb{E}U_1\,(I\{U_1 \geq U\} - \tfrac{1}{2}) \;=\; \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12}.\end{aligned}$$

By symmetry $\mathbb{E}\, I\{X_1 \geq X_2\}\, \xi_2 = \mathbb{E}\xi_2 - \mathbb{E}\, I\{X_2 \geq X_1\}\, \xi_2 = 0 - \frac{1}{12} = -\frac{1}{12}$ and we see that $b = -\frac{2}{12} - \frac{2}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{4} = \frac{1}{12}$. As a result

$$\Delta_1^2 \;=\; \tfrac{1}{12}\, n^2\, \rho_2, \tag{4.8}$$

and Theorem 4.1 is an easy consequence of Theorem 3.3, (4.4), (4.6), (4.7) and (4.8). □

We have the following nice application of Theorem 4.1. Let $Y_1, \ldots, Y_{n_1}$ and $Z_1, \ldots, Z_{n_2}$ be independent i.i.d. samples, with continuous distribution functions $F_Y$ and $F_Z$. We may test the null hypothesis that they are equal by using Wilcoxon's rank-sum statistic, defined by

$$\mathbb{T} \;:=\; \textstyle\sum_{j=1}^{n_1} R_j, \tag{4.9}$$

that is, $\mathbb{T}$ of the form (4.3), taking $a_j := I\{1 \leq j \leq n_1\}$ for $j = 1, \ldots, n$, writing as well $n := n_1 + n_2$. Here

$$\rho_2 \;=\; n^{-1}\left(n_1\,(1 - \tfrac{n_1}{n})^2 + n_2\,(-\tfrac{n_1}{n})^2\right) \;= n^{-3}\, n_1\, n_2\,(n_1 + n_2) \;=\; n^{-2}\, n_1\, n_2$$

and

$$\rho_3 \;=\; n^{-1}\left(n_1\,|1 - \tfrac{n_1}{n}|^3 + n_2\,| - \tfrac{n_1}{n}|^3\right) \;= n^{-4}\, n_1\, n_2\,(n_1^2 + n_2^2),$$

so that

$$\rho_2^{-3/2}\,\rho_3 \;=\; n^{-1}\, n_1^{-1/2}\, n_2^{-1/2}\,(n_1^2 + n_2^2) \;\leq\; (n_1/n_2)^{1/2} + (n_2/n_1)^{1/2}.$$

A straightforward application of Theorem 4.1 now brings us the following:

**Corollary 4.2.** *Let $\mathbb{T}$ be Wilcoxon's rank-sum statistic, as defined in (4.9). Under the null hypothesis that $F = G$ and that $F$ and $G$ are continuous, an absolute constant $c$ exists such that*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left( \frac{\mathbb{T} - \frac{1}{2}n_1\,(n+1)}{\{\frac{1}{12}n_1\, n_2\, n\}^{1/2}} \leq x \right) - \Phi(x) \right| \;\leq\; c\, n^{-1/2} \left\{ \left(\frac{n_1}{n_2}\right)^{1/2} + \left(\frac{n_2}{n_1}\right)^{1/2} \right\}.$$

Notice that, following (4.1), we easily derive the concentration bound

$$Q(\mathbb{T}/s,\lambda) \ \ll\ \max\{\lambda, \rho_2^{-3/2}\,\rho_3\,n^{-1/2}\}$$

in the general case, and $Q(\mathbb{T}/s,\lambda) \ll \max\{\lambda, n^{-1/2}\,\big((n_1/n_2)^{1/2}+(n_2/n_1)^{1/2}\big)\}$ in the case of Wilcoxon's rank-sum statistic.

## 4.2 $U$-statistics

Next we take a look at $U$-statistics of order $k$, for $k \geq 2$. It does not make much sense to look at non-i.i.d. samples of changing kernels at the moment, since in these cases the results of Theorem 2.2 and Theorem 3.3 do not really simplify. Because of this, we confine ourselves to the case of an i.i.d. sample $X_1,\dots,X_n$ and look at the statistic

$$\mathbb{T} \ :=\ \textstyle\sum_{1\leq j_1<\cdots<j_k\leq n} h(X_{j_1},\dots,X_{j_k}), \tag{4.10}$$

where $h:\mathcal{X}^k \to \mathbb{R}$ is some symmetric Borel function. We suppose that

$$\mathbb{E}\,h := \mathbb{E}\,h(X_1,\dots,X_k) = 0 \quad\text{and}\quad \mathbb{E}\,h^2 := \mathbb{E}\,h^2(X_1,\dots,X_k) < \infty,$$

that is to say, that $\mathbb{E}\,\mathbb{T} = 0$ and $\sigma^2 := \operatorname{var}\mathbb{T} < \infty$. For $l = 1,\dots,k$, we write

$$g_l(X_1,\dots,X_l) \ :=\ [h(X_1,\dots,X_k)]_{1,\dots,l}, \tag{4.11}$$

as defined in (A.4), that is,

$$g_l(X_1,\dots,X_l) \ =\ \textstyle\sum_{B\subset\{1,\dots,l\}} (-1)^{l-|B|}\,\mathbb{E}\,(h(X_1,\dots,X_k)\,|\,B).$$

The functions $g_l$ are (almost surely) uniquely determined. For $p = 2,3$ and $l = 2,\dots,k$, we define the moments

$$\xi_p \ :=\ \mathbb{E}\,|g_1(X_1)|^p \quad\text{and}\quad \theta_{2,l} \ :=\ \mathbb{E}\,|g_l(X_1,\dots,X_l)|^2. \tag{4.12}$$

By construction of the Hoeffding decomposition now

$$\begin{aligned} \mathbb{E}\,h^2 \ &=\ \textstyle\sum_{l=1}^k \binom{k}{l}\,\mathbb{E}\,|g_l(X_1,\dots,X_l)|^2 \\ &=\ k\,\xi_2 + \binom{k}{2}\,\theta_{2,2} + \textstyle\sum_{l=3}^k \binom{k}{l}\,\theta_{2,l} \ =:\ \alpha_1+\alpha_2+\alpha_3. \end{aligned} \tag{4.13}$$

We follow the lines of Van Zwet (1984), near Corollary 4.1. Here the following Berry-Esseen type result for $\mathbb{T}$ is derived: there exists an absolute constant $c$ such that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T}/\sigma \le x) - \Phi(x)| \;\le\; c\,n^{-1/2}\,(\xi_2^{-3/2}\,\xi_3 + k^2\,\xi_2^{-1}\,\mathbb{E}\,h^2). \qquad (4.14)$$

Application of Theorem 2.2 and Theorem 3.3 leads us to the following concentration and Berry-Esseen bound:

**Corollary 4.3.** *Assume that the sample $X_1, \ldots, X_n$ is i.i.d. and that $\mathbb{T}$ is of the form (4.10). Moreover, assume that $\mathbb{E}\,\mathbb{T} = 0$, $\xi_2 > 0$ and $\mathbb{E}\,h^2 < \infty$. Then $s^2 = n\binom{n-1}{k-1}^2 \xi_2$, and there exists an absolute constant $c$ such that*

$$Q(\mathbb{T}/s, \lambda) \;\le\; c\,\max\big\{\lambda\,,\,\xi_2^{-3/2}\,\xi_3\;n^{-1/2} + (\xi_2^{-1}\,\mathbb{E}\,h^2)\,(1+k^3 n^{-1})\;n^{-1}\big\}, \qquad (4.15)$$

*and, with $D$ as in* (4.1),

$$D \;\le\; c\;n^{-1/2}\,\big(\xi_2^{-3/2}\,\xi_3 + \big(\xi_2^{-1}\,\mathbb{E}\,h^2\;(1+k^3 n^{-1})\big)^{1/2}\big). \qquad (4.16)$$

*We have the same results with $\sigma$ instead of $s$.*

Remark: as $\xi_2^{-1}\,\mathbb{E}\,h^2 \ge 1$, the Berry-Esseen bound (4.16) is sharper than (4.14).

**Proof of Corollary 4.3.** As promised we will use Theorem 2.2 and Theorem 3.3. In order to do this, we need to determine $s^2$, $\beta$, $\Delta_1^2$ and $\Delta_2^2$.

Following (A.5), for any $A \subset N$ we have that

$$T_A \;=\; \textstyle\sum_{1\le j_1<\cdots<j_k\le n} [h(X_{j_1}, \ldots, X_{j_k})]_A.$$

As to $[h(X_{j_1}, \ldots, X_{j_k})]_A$, for any $A = \{n_1, \ldots, n_l\}$ (with $1 \le l \le n$), suppose that $n_p \notin \{j_1, \ldots, j_k\}$. With (A.10) then $D_{n_p} h(X_{j_1}, \ldots, X_{j_k}) = h(X_{j_1}, \ldots, X_{j_k}) D_{n_p} 1 = 0$, so that with (A.11) we see that

$$\begin{aligned}[h(X_{j_1}, \ldots, X_{j_k})]_A \;&=\; \mathbb{E}\,(D_A h(X_{j_1}, \ldots, X_{j_k}) \,|\, A) \\ &=\; \mathbb{E}\,(D_{A\setminus\{n_p\}} D_{n_p} h(X_{j_1}, \ldots, X_{j_k}) \,|\, A) \;=\; 0.\end{aligned}$$

In the other case we have that $A \subset \{j_1, \ldots, j_k\}$, and using symmetry we see that then

$$[h(X_{j_1}, \ldots, X_{j_k})]_A \;=\; g_l(X_{n_1}, \ldots, X_{n_l}). \qquad (4.17)$$

As a consequence of this, for $l = 1, \ldots, n$,

$$T_{1,\ldots,l} \;=\; \tbinom{n-l}{k-l}\, g_l(X_1, \ldots, X_l). \tag{4.18}$$

Using (4.18) we see that

$$s^2 = n\,\mathbb{E}\,T_1^2 = n\,\tbinom{n-1}{k-1}^2\,\xi_2 \quad \text{and} \quad \beta = n\,\mathbb{E}\,|T_1|^3 = n\,\tbinom{n-1}{k-1}^3\,\xi_3, \tag{4.19}$$

so that, by assumption, $s^2 > 0$, and $\beta\, s^{-3} = n^{-1/2}\,\xi_2^{-3/2}\,\xi_3$. Moreover, (4.18) leads us to the fact that $\Delta_1^2 = \binom{n}{2}\,\mathbb{E}\,T_{1,2}^2 = \binom{n}{2}\,\binom{n-2}{k-2}^2\,\theta_{2,2}$, so that, cf. (4.13),

$$\begin{aligned} \Delta_1^2\, s^{-2} \;&=\; \tbinom{n}{2}\, n^{-1}\,\tbinom{n-2}{k-2}^2\,\tbinom{n-1}{k-1}^{-2}\,\xi_2^{-1}\,\theta_{2,2} \\ &\le\; \tfrac{1}{2} n\,\left(\tfrac{k-1}{n-1}\right)^2 \xi_2^{-1}\,\theta_{2,2} \;\ll\; k^2\, n^{-1}\,\xi_2^{-1}\,\theta_{2,2} \;\le\; n^{-1}\,\xi_2^{-1}\,\mathbb{E}\,h^2. \end{aligned} \tag{4.20}$$

Finally we look at $\Delta_2^2$. In case $k = 2$ we have that $\Delta_2^2 = 0$, so we suppose that $k \ge 3$. Using (A.18) we see that

$$\Delta_2^2 \;=\; \textstyle\sum_{l=3}^{k} \binom{l}{2}\binom{n}{l}\,\mathbb{E}\,T_{1,\ldots,l}^2 \;=\; \sum_{l=3}^{k} \binom{l}{2}\binom{n}{l}\binom{n-l}{k-l}^2\,\theta_{2,l}.$$

Here in general $\binom{l}{2}\binom{n}{l} \;=\; \binom{n}{2}\binom{n-2}{l-2}$ and $\binom{n-2}{l-2}\binom{n-l}{k-l}^2 \;=\; \frac{l(l-1)}{n(n-1)}\,\binom{n}{k}\binom{n-l}{k-l}\binom{k}{l}$, which leads us to

$$\Delta_2^2 \;\le\; \tfrac{k(k-1)}{n(n-1)}\,\tbinom{n}{2}\tbinom{n}{k}\,\textstyle\sum_{l=3}^{k}\binom{n-l}{k-l}\binom{k}{l}\,\theta_{2,l} \;\le\; \tfrac{1}{2}k^2\,\binom{n}{k}\binom{n-3}{k-3}\,\alpha_3.$$

Hence

$$\Delta_2^2\, s^{-2} \;\ll\; k^2\, n^{-1}\,\tbinom{n}{k}\tbinom{n-3}{k-3}\tbinom{n-1}{k-1}^{-2}\,\xi_2^{-1}\,\alpha_3,$$

and, since in general $\binom{n}{k}\binom{n-3}{k-3} \le k\, n^{-1}\,\binom{n-1}{k-1}^2$,

$$\Delta_2^2\, s^{-2} \;\ll\; k^3\, n^{-2}\,\xi_2^{-1}\,\alpha_3 \;\le\; k^3\, n^{-2}\,\xi_2^{-1}\,\mathbb{E}\,h^2. \tag{4.21}$$

Note that $\sigma^2 \le s^2 + \Delta_1^2 + \Delta_2^2 < \infty$ by the above. Using Theorem 3.3 together with (4.19), (4.20) and (4.21), the statement in (4.16) follows from this.

In the same way it follows via Theorem 2.2 that the concentration bound (4.15) is correct. This finishes the proof. $\square$

## 4.3 Incomplete $U$-statistics

Next we have a look at so-called incomplete $U$-statistics. Here the idea is that it is often easier to, instead of looking at regular $U$-statistics, take into account only a limited amount of summands. This will save time and effort, while it does not have to lead to significant loss of information.

Let again $k \geq 2$, let $X_1, \ldots, X_n : \Omega \to \mathcal{X}$ be an i.i.d. sample and suppose that $h : \mathcal{X}^k \to \mathbb{R}$ is a symmetric Borel function. In the previous section we have been considering $U$-statistics of order $k$, which were of the form (4.10). Incomplete $U$-statistics on the other hand are of the form

$$U \ := \ \textstyle\sum_{J \in C} \gamma_J \, h(X_J), \tag{4.22}$$

using the set

$$C \ := \ \{(j_1, \ldots, j_k) \in N^k \ : \ 1 \leq j_1 < \cdots < j_k \leq n\},$$

and the expression $h(X_J) = h(X_{j_1}, \ldots, X_{j_k})$ in case $J = (j_1, \ldots, j_k)$. The coefficients $\gamma_J$ are either stochastic or deterministic, and are expected to contain a substantial amount of 0's, whereas in other cases they are usually equal to 1.

From here on, let $K := \binom{n}{k}$, let $1 \leq M = M(n) \leq K$ be some constant, representing the (expected) sum of the $\gamma_J$, and let

$$\gamma \ := \ (\gamma_J : J \in C)$$

be the stochastic $K$-vector associated with the $\gamma_J$. In the stochastic scheme we assume that $\gamma$ is independent of the sample and that we are in one of the following three cases:

(1) The $\gamma_J$ are independent, sharing a Bernoulli-distribution with parameter $MK^{-1}$. That is to say, for all $J$, $\mathbb{P}(\gamma_J = 1) = 1 - \mathbb{P}(\gamma_J = 0) = MK^{-1}$.

(2) The stochastic $K$-vector $\gamma$ has a multinomial distribution with parameters $M$ and $K^{-1}$ (sampling with replacement). That is to say, we have that

$$\mathbb{P}(\gamma = \mathbf{l}) \ = \ \tfrac{M!}{l_1! \cdots l_K!} \, K^{-M}$$

for all combinations $\mathbf{l} = (l_1, \ldots, l_K)$ of non-negative integers such that $\sum_{j=1}^{K} l_j = M$.

(3) Sampling without replacement: $M$ of the $\gamma_J$ equal 1 and the others equal 0, the 1's being uniformly distributed over the $\gamma_J$. That is to say, $\mathbb{P}(\gamma = \mathbf{l}) = \binom{K}{M}^{-1}$ for any $\mathbf{l} = (l_1, \ldots, l_K)$ such that $l_j \in \{0,1\}$ and $\sum_{j=1}^K l_j = M$.

For $M$ the smaller is the better, as this will save the most time and effort. Of course, taking $M$ too small might mean getting rid of too much information, so next to optimal concentration and Berry-Esseen bounds, we are looking as well for an optimal size of $M$.

A rigorous exposition of the above may be found in Blom (1976) and Janson (1984). In these papers central limit theorems are derived for $U$, under varying conditions. We confine ourselves to the case in which $n/M \to 0$ and look for Berry-Esseen type results. We start by dealing with the stochastic case, writing $g_l$, $\xi_p$ and $\theta_{2,l}$ as in (4.11) and (4.12).

We have the following:

**Theorem 4.4.** *Let $X_1, \ldots, X_n$ be an i.i.d. sample and let $U$ be of the form* (4.22)*, the $\gamma_J$ being of the form* (1), (2) *or* (3). *Assume that $\mathbb{E}\,h = 0$, $\xi_2 > 0$ and $\mathbb{E}\,h^2 < \infty$. Then $s^2 = s^2(U) = k^2\,M^2\,n^{-1}\,\xi_2$, and, using the notation $D = D(U)$ as in* (4.1)*, there exists an absolute constant $c$ such that*

$$D \;\leq\; c\,n^{-1/2}\left(\xi_2^{-3/2}\,\xi_3 + (\xi_2^{-1}\,\mathbb{E}\,h^2)^{1/2}\,(1 + k^3 n^{-1} + n^2/M)^{1/2}\right). \tag{4.23}$$

*Under the same assumptions we have the following concentration bound:*

$$\begin{aligned} Q(U/s, \lambda) \;\;\leq\;\; c\;\max\bigl\{\lambda\,,\,(\xi_2^{-3/2}\,\xi_3)\,n^{-1/2} \\ +\;(\xi_2^{-1}\,\mathbb{E}\,h^2)\,(1 + k^3 n^{-1} + n^2/M)\,n^{-1}\bigr\}. \end{aligned} \tag{4.24}$$

*We have the same results with $\sigma$ instead of $s$.*

Theorem 4.4 provides a direct generalization of Corollary 4.3, which is easily seen by taking $M = K$. If $k$ is fixed for $M, n \to \infty$, the moments $\xi_2$, $\xi_3$ and $\mathbb{E}\,h^2$ are fixed as well, that is to say, only depending on the distribution of the sample. In that case, if $n^2\,M^{-1} \ll 1$, (4.23) gives us a Berry-Esseen bound for $U$. Taking $M$ much larger than $n^2$ will not improve our bound any further, and therefore seems unsensible. Under the weaker condition that $n/M \to 0$, (4.23) gives us weak convergence of $U/s$ to the standard normal distribution.

In the case that *not* $n/M \to 0$, (4.23) comes down to the trivial bound $D \ll 1$, in which case Theorem 4.4 is of no use. This is due to the fact that in our set-up we are concentrating essentially on situations where the linear part of the statistic is dominating, that is to say, where $s^2 \approx \sigma^2$. If, say, $\mathrm{var}(\mathbb{T}_2 + \mathbf{T}_3) \geq \frac{1}{3}\sigma^2$, it will follow that $\Delta^2 s^{-2} \geq \frac{1}{3}$, and the resulting bound is clearly useless. In the present case, following (4.31), (4.26) and (4.27), we have that $s^2$ is of order $M^2 n^{-1}$, whereas $\mathrm{var}(\mathbb{T}_2 + \mathbf{T}_3)$ is of order $M^2 n^{-2} + M$, so that the linear part is precisely dominating if $M^2 n^{-1}$ is really larger than $M$, which corresponds to our original demand that $n/M \to 0$. In fact, following a different method, Janson (1984) obtains a central limit theorem even in the case where $M/n \to 0$.

**Proof of Theorem 4.4.** First we gather some basic facts concerning the $\gamma_J$. First, let from here on

$$\mathbb{E}\gamma \;:=\; MK^{-1}.$$

For all $I, J \in C$ with $I \neq J$ we have that

$$\mathbb{E}\gamma_J \;=\; \mathbb{E}\gamma, \quad \mathrm{var}\,\gamma_J \;\leq\; MK^{-1} \quad \text{and} \quad \mathrm{cov}(\gamma_I, \gamma_J) \;\leq\; 0. \tag{4.25}$$

Indeed, the expectation values are easily checked. As to the variance: in the cases (1) and (3) we have that $\mathrm{var}\,\gamma_J = MK^{-1}(1 - MK^{-1})$, whereas in case (2) the variance equals $MK^{-1}(1 - K^{-1})$. As to the covariance: in case (1) it clearly equals 0 (because of independence), whereas in the cases (2) and (3) it is intuitively clear that if $\gamma_I$ is for example 1 this has a negative influence on the chance of $\gamma_J$'s being bigger than 0. This argument is easily formalized.

Now we just consider the statistic

$$U \;=\; U(X_1, \ldots, X_n, \gamma)$$

as a function of $n+1$ independent random variables and determine its Hoeffding decomposition. In fact, taking any subset $\emptyset \neq A \subset N$ such that $|A| \leq k$, we have that

$$U_A \;=\; U_A(X_A) \;=\; (\mathbb{E}\gamma)\left[\textstyle\sum_{J \in C} h(X_J)\right]_A \;=\; M \binom{n}{k}^{-1} \binom{n-|A|}{k-|A|} g_{|A|}(X_A) \tag{4.26}$$

and

$$\begin{aligned} U_{A,\gamma} \;=\; U_{A,\gamma}(X_A, \gamma) \;&=\; \textstyle\sum_{J: A \subset J} (\gamma_J - \mathbb{E}\gamma)\,[h(X_J)]_A \\ &=\; g_{|A|}(X_A) \textstyle\sum_{J: A \subset J} (\gamma_J - \mathbb{E}\gamma), \end{aligned} \tag{4.27}$$

and all other terms of the Hoeffding decomposition vanish.

We first prove that (4.26) and (4.27) describe $U$'s Hoeffding decomposition. Indeed, taking $A \subset N$, using Lemma A.2 and (A.10), we have that

$$\begin{aligned} U_A &= \mathbb{E}(D_A U \mid A) \;=\; \mathbb{E}(\mathbb{E}(D_A U \mid N) \mid A) \\ &= \mathbb{E}(\textstyle\sum_{J\in C} \mathbb{E}(\gamma_J\, D_A h(X_J) \mid N)) \mid A) \;=\; \mathbb{E}(\textstyle\sum_{J\in C} D_A h(X_J)\, \mathbb{E}\gamma \mid A) \\ &= (\mathbb{E}\gamma)\, \mathbb{E}(D_A \textstyle\sum_{J\in C} h(X_J) \mid A) \;=\; (\mathbb{E}\gamma)\, [\textstyle\sum_{J\in C} h(X_J)]_A. \end{aligned} \tag{4.28}$$

As in (4.17), here

$$[h(X_J)]_A \;=\; g_{|A|}(X_A)\, I\{A \subset J\}, \tag{4.29}$$

and (4.26) follows easily. On the other hand we have that

$$U_{A,\gamma} \;=\; \textstyle\sum_{J\in C} [\gamma_J\, h(X_J)]_{A,\gamma},$$

whereas, using again Lemma A.2 and (A.10),

$$\begin{aligned} [\gamma_J\, h(X_J)]_{A,\gamma} &= \mathbb{E}(D_{A,\gamma} \gamma_J\, h(X_J) \mid A, \gamma) \\ &= \mathbb{E}(D_\gamma \gamma_J\, D_A h(X_J) \mid A, \gamma) \;=\; \mathbb{E}((D_A h(X_J))\,(D_\gamma \gamma_J) \mid A, \gamma) \\ &= \mathbb{E}(D_\gamma \gamma_J \mid A, \gamma)\, \mathbb{E}(D_A h(X_J) \mid A, \gamma) \\ &= (\gamma_J - \mathbb{E}\gamma)\, \mathbb{E}(D_A h(X_J) \mid A) \;=\; (\gamma_J - \mathbb{E}\gamma)\, [h(X_J)]_A, \end{aligned}$$

and from (4.29) it follows that (4.27) is correct. It is easily seen that the other terms vanish.

We may calculate the appropiate moments now and apply Theorem 3.3, but the $\varepsilon_2$ over there spoils our bound a bit. Using the methods that we used earlier to get rid of $\varepsilon_2$ in the i.i.d. symmetric case, we may remove it here too. Thus, we follow again the non-i.i.d. proof of Section 3.5 in order to show that in this special case

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(U/s \le x) - H(x)| \;\le\; c(\varepsilon_1)\, \max\left\{\beta\, s^{-3},\; (\Delta_1^2\, s^{-2})^{1-\varepsilon_1},\; (\Delta_2^2\, s^{-2})^{1/2}\right\} \tag{4.30}$$

for some constant $c(\varepsilon_1)$, cf. (3.9).

In fact, as before we work under the assumptions that $\mathbb{E}U = 0$ and $s^2 = 1$, thus really looking at the statistic $\mathbb{T} := U/s$. Taking $\varepsilon := \varepsilon_1$ and

$$a \;:=\; \min\{\beta^{-1}, (\Delta_1^2)^{-(1-\varepsilon)}, (\Delta_2^2)^{-1/2}\} \;\ge\; 100$$

we prove that

$$\int_0^a |\hat{f}(t) - \hat{h}(t)|\, t^{-1}\, dt \;\ll\; a^{-1}.$$

In fact, to this, we need to improve our results a bit on $[0, a^{1/(2k-2)}]$ and $[a^{2^{-(l+1)}}, a^{2^{-l}}]$, whereas with (3.70) the interval $[a^{1/2}, a]$ can be treated as before.

As to the interval $[0, a^{1/(2k-2)}]$, for the removal of $\mathbf{T}_3$ we need to use a Taylor expansion which is one term longer, which leaves us the estimation of the two integrals

$$\delta_1 \;:=\; \int_0^{a^{1/(2k-2)}} |\mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\}\, \mathbf{T}_3|\, dt \quad \text{and} \quad \delta_2 \;:=\; \tfrac{1}{2} \int_0^{a^{1/(2k-2)}} t\, \mathbb{E}\, \mathbf{T}_3^2\, dt.$$

Clearly $\delta_2 \le \frac{1}{4} \Delta_2^2\, a^{1/(k-1)} \ll a^{-1}$ and

$$\delta_1 \;\le\; \delta_3 + \delta_4 \;:=\; \int_0^{a^{1/(2k-2)}} |\mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbf{T}_3|\, dt + \int_0^{a^{1/(2k-2)}} t\, \mathbb{E}\, |\mathbb{T}_2|\, |\mathbf{T}_3|\, dt,$$

with $\delta_4 \ll a^{-1}$ as in (3.35). As to $\delta_3$ we have that

$$\mathbf{T}_3 \;=\; S_1 + S_2 \;:=\; \sum\nolimits_{A \subset N:\, |A| \ge 3} T_A + \sum\nolimits_{A \subset N:\, |A| \ge 2} T_{A,\gamma},$$

with

$$\mathbb{E}\, e\{t\mathbb{T}_1\}\, S_2 \;=\; \mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbb{E}\, (S_2 \mid N) \;=\; 0,$$

so that $\delta_3 = \int_0^{a^{1/(2k-2)}} |\mathbb{E}\, e\{t\mathbb{T}_1\}\, S_1|\, dt$, for which the method of Lemma 3.8 is applicable. As in (3.39), with $\gamma(t)$ and $\theta(t)$ as in (3.36), we see that

$$\begin{aligned} |\mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbf{T}_3| \;&\ll\; (1 - \theta(t))^{-1} \left( \sum\nolimits_{l=3}^k \tbinom{n-2}{l-2}\, \mathbb{E}\, T_{1,\dots,l}^2 \right)^{1/2} \\ &\le\; (1 - \theta(t))^{-1} \left( n^{-2}\, (\tilde{\delta}_2\, n^{-1}) \right)^{1/2} \\ &\ll\; n^{-1}\, (1 - \theta(t))^{-1}\, (\Delta_2^2)^{1/2}, \end{aligned}$$

see as well (A.22), (A.25) and (A.26). Here as in (3.43)

$$1 - \theta(t) \;=\; n^{-1}\, t^2 - R \;\ge\; n^{-1}\, t^2\, (1 - \tfrac{1}{6}\beta^{-1} \cdot 4\beta) \;=\; \tfrac{1}{3} n^{-1}\, t^2,$$

so that $|\mathbb{E}\, e\{t\mathbb{T}_1\}\, \mathbf{T}_3| \ll t^{-2}\, a^{-1}$ and, as in (3.44), $\delta_3 \ll a^{-1}$.

On the intervals of the form $[a^{2^{-(l+1)}}, a^{2^{-l}}]$ the same approach is working, see as well the proof of Lemma 3.9, and we see that indeed

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\mathbb{T} \le x) - H(\mathbb{T}, x)| \;\le\; c(\varepsilon_1)\, \max \left\{ \beta(\mathbb{T}),\; \Delta_1^2(\mathbb{T})^{1-\varepsilon_1},\; \Delta_2^2(\mathbb{T})^{1/2} \right\}$$

for some constant $c(\varepsilon_1)$. As $\mathbb{T} = U/s$ this is in fact telling us that we have (4.30), and all we need to do now is collect the moments.

The moments are easily calculated using (4.26) and (4.27). In fact we have that

$$s^2 \;=\; n\,\mathbb{E}U_1^2 \;=\; n\,M^2\binom{n}{k}^{-2}\binom{n-1}{k-1}^2\xi_2 \;=\; k^2\,M^2\,n^{-1}\,\xi_2, \tag{4.31}$$

$$\beta \;=\; n\,\mathbb{E}|U_1|^3 \;=\; n\,M^3\binom{n}{k}^{-3}\binom{n-1}{k-1}^3\xi_3 \;=\; k^3\,M^3\,n^{-2}\,\xi_3$$

and

$$\begin{aligned}\Delta_1^2 \;&=\; \binom{n}{2}M^2\binom{n}{k}^{-2}\binom{n-2}{k-2}^2\theta_{2,2} + n\,\xi_2\,\mathbb{E}|\textstyle\sum_{1\in J}(\gamma_J - \mathbb{E}\gamma)|^2\\ &=\; \tfrac{k^2\,(k-1)^2\,M^2}{2n(n-1)}\,\theta_{2,2} + n\,\xi_2\,\mathrm{var}\,\textstyle\sum_{1\in J}\gamma_J.\end{aligned}$$

Here because of symmetry and (4.25)

$$\begin{aligned}\mathrm{var}\,\textstyle\sum_{1\in J}\gamma_J \;&=\; \textstyle\sum_{1\in J}\mathrm{var}\,\gamma_J + \sum_{1\in I,J,\;I\neq J}\mathrm{cov}(\gamma_I,\gamma_J)\\ &\le\; \binom{n-1}{k-1}M\binom{n}{k}^{-1} \;=\; \tfrac{k}{n}\,M,\end{aligned} \tag{4.32}$$

so that

$$\Delta_1^2 \;\ll\; k^4\,M^2\,n^{-2}\,\theta_{2,2} + k\,M\,\xi_2.$$

Using the same type of argument we see that (see (A.19))

$$\begin{aligned}\Delta_2^2 \;&=\; \textstyle\sum_{l=3}^{k}\binom{l}{2}\binom{n}{l}\mathbb{E}U_{1,\dots,l}^2 + \sum_{l=3}^{k+1}\binom{l}{2}\binom{n}{l-1}\mathbb{E}U_{1,\dots,l-1,\gamma}^2\\ &=\; \textstyle\sum_{l=3}^{k}\binom{l}{2}c_l\,\theta_{2,l} + \sum_{l=3}^{k+1}\binom{l}{2}d_{l-1}\,\theta_{2,l-1},\end{aligned}$$

with

$$c_l \;:=\; \binom{n}{l}M^2\binom{n}{k}^{-2}\binom{n-l}{k-l}^2$$

and

$$d_{l-1} \;:=\; \binom{n}{l-1}\mathbb{E}|\textstyle\sum_{J:\{1,\dots,l-1\}\subset J}(\gamma_J - \mathbb{E}\gamma)|^2.$$

Here we have that

$$\begin{aligned}c_l \;&=\; M^2\,\tfrac{n!}{l!\,(n-l)!}\,\tfrac{k!^2\,(n-k)!^2}{n!^2}\,\tfrac{(n-l)!^2}{(k-l)!^2\,(n-k)!^2}\\ &=\; M^2\binom{k}{l}\tfrac{k(k-1)\cdots(k-l+1)}{n(n-1)\cdots(n-l+1)} \;\ll\; M\binom{k}{l}M\,k^3\,n^{-3}\end{aligned}$$

and, as in (4.32),

$$\begin{aligned} d_l &= \binom{n}{l} \operatorname{var} \textstyle\sum_{J:\{1,\dots,l\}\subset J} \gamma_J \;\leq\; \binom{n}{l}\binom{n-l}{k-l} \operatorname{var}\gamma_J \\ &\leq M \frac{n!}{l!\,(n-l)!} \frac{(n-l)!}{(k-l)!\,(n-k)!} \frac{k!\,(n-k)!}{n!} \;=\; M\binom{k}{l}, \end{aligned}$$

so that (see (4.13))

$$\Delta_2^2 \;\ll\; k^2 (1 + Mk^3 n^{-3})\, M \textstyle\sum_{l=2}^{k} \binom{k}{l} \theta_{2,l} \;\ll\; k^2 (1 + Mk^3 n^{-3})\, M\, \mathbb{E} h^2.$$

As a result we see that

$$\beta\, s^{-3} \;=\; \xi_3\, \xi_2^{-3/2}\, n^{-1/2}, \qquad \Delta_1^2\, s^{-2} \;\ll\; k^2\, \theta_{2,2}\, \xi_2^{-1}\, n^{-1} + k^{-1}\, n\, M^{-1}$$

and

$$\Delta_2^2\, s^{-2} \;\ll\; n\, M^{-1} (1 + M\, k^3\, n^{-3})\; \xi_2^{-1}\, \mathbb{E} h^2 \;=\; (n\, M^{-1} + k^3\, n^{-2})\; \xi_2^{-1}\, \mathbb{E} h^2.$$

Using (4.30) for $\varepsilon_1 = \frac{1}{2}$, we now easily see that (4.23) is correct. As to the concentration bound we just apply Theorem 2.2 and obtain (4.24). □

In the case where the $\gamma_J$ are fixed, as in the proof of Theorem 4.4 it appears that

$$U \;=\; U(X_1, \dots, X_n) \;=\; \textstyle\sum_{A\subset N} U_A$$

with

$$\begin{aligned} U_A &= \textstyle\sum_{J\in C} \gamma_J\, \mathbb{E}(D_A\, h(X_J) \,|\, A) \;=\; \textstyle\sum_{J\in C} \gamma_J\, [h(X_J)]_A \\ &= \textstyle\sum_{J\in C} \gamma_J\; g_{|A|}(X_A)\, I\{A\subset J\} \;=\; g_{|A|}(X_A) \textstyle\sum_{J:A\subset J} \gamma_J, \end{aligned}$$

see (4.28) and (4.29). As a result

$$s^2 = \xi_2 \textstyle\sum_{j=1}^{n} (\sum_{j\in J} \gamma_J)^2, \qquad \beta = \xi_3 \textstyle\sum_{j=1}^{n} (\sum_{j\in J} \gamma_J)^3,$$

$$\Delta_1^2 = \theta_{2,2} \textstyle\sum_{A:|A|=2} (\sum_{J:A\subset J} \gamma_J)^2$$

and

$$\Delta_2^2 \;=\; \textstyle\sum_{l=3}^{k} \binom{l}{2}\, \theta_{2,l} \sum_{A:|A|=l} \big(\sum_{J:A\subset J} \gamma_J\big)^2,$$

and, using appropriate $\gamma_J$, Theorems 2.2 and 3.3 again lead to a concentration and Berry-Esseen bound.

## 4.4 Self-normalized statistics: concentration bounds

Next we take a look at so-called self-normalized statistics. What we are in fact thinking of are Student's statistic in a one- and a two-sample setting and two standard statistics associated with the model of linear regression. These will be considered in the following two subsections. The four statistics have a common structure that is asking for a unified approach. The price we pay for this will be that our key results, Theorems 4.5 and 4.10, are of a rather abstract form.

Let $X_1, \ldots, X_n$ be independent, not necessarily identically distributed real-valued random variables. Writing $X := (X_1, \ldots, X_n)$, let $R = R(X)$ be a (supposedly negligible) random variable such that, for all $\lambda$,

$$R(\lambda X) \;=\; \lambda^2 R(X). \tag{4.33}$$

We are typically thinking of cases like $R(X) = 0$ or $R(X) = -n\,\overline{X}^2$, with $\overline{X}$ denoting the sample mean. Let $c_1, \ldots, c_n$ moreover be fixed real numbers, for which we will assume that $\sum_{j=1}^n c_j^2 = n$. We confine ourselves to the following two cases:

1. The sample $X_1, \ldots, X_n$ is identically distributed,
2. All the coefficients $c_j$ are equal to 1.

An application of Theorem 2.2 may be found for the general self-normalized statistic

$$\mathbb{T} \;:=\; \frac{\sum_{j=1}^n c_j X_j}{\big(\sum_{k=1}^n X_k^2 + R(X)\big)^{1/2}}. \tag{4.34}$$

Let $M_j \ge 0$, for $1 \le j \le n$, denote fixed positive numbers, usually depending on the probability distribution of $X_j$, and let

$$s^2 \;:=\; \textstyle\sum_{j=1}^n \mathbb{E}\, X_j^2\, I\{X_j^2 \le M_j^2\}, \tag{4.35}$$

$$U_j \;:=\; s^{-1} X_j\, I\{X_j^2 \le M_j^2\} \qquad \text{and} \qquad s_j^2 \;:=\; \mathbb{E}\, U_j^2, \tag{4.36}$$

assuming throughout that $s^2 \neq 0$. We write $U := (U_1, \dots, U_n)$. We will assume as well that

$$M_j \;\leq\; s$$

for all $j$, so that by construction $|U_j| \leq 1$ and $\sum_{j=1}^n s_j^2 = 1$. We write

$$A_1 \;:=\; \sum_{j=1}^n \mathbb{P}(X_j^2 > M_j^2), \quad A_2 \;:=\; \sum_{j=1}^n |\mathbb{E}U_j|, \quad A_3 \;:=\; \sum_{j=1}^n \mathbb{E}|U_j|^3. \tag{4.37}$$

Taking a random variable $\alpha$ such that $\mathbb{P}(\alpha = s_j^2) = \frac{1}{n}$ for $1 \leq j \leq n$, and using Lyapunov's inequality, we see that

$$A_3 \;\geq\; \sum_{j=1}^n (s_j^2)^{3/2} \;=\; n\,\mathbb{E}|\alpha|^{3/2} \;\geq\; n\,(\mathbb{E}|\alpha|)^{3/2} \;=\; n^{-1/2}. \tag{4.38}$$

For $p > 0$, we write

$$\nu_p \;:=\; n^{-1} \sum_{j=1}^n |c_j|^p.$$

Note that $\nu_2 = 1$, so with Lyapunov's inequality $\nu_1 \leq 1 \leq \nu_3$. In case $c_j = 1$ for all $j$, we have that $\nu_p = 1$ for any $p$.

As concerns the concentration bound of $\mathbb{T}$, we have the following general result:

**Theorem 4.5.** *Assume that $s^2 \neq 0$, $M_j \leq s$ for all $j$ and $\sum_{j=1}^n c_j^2 = n$. Moreover, assume that either the sample $X_1, \dots, X_n$ is i.i.d. or that $c_j = 1$ for all $j$. Finally assume that (4.33) applies and $\mathbb{E}|R(U)|^{3/2} \ll A_3^3$. Then there exists an absolute constant $c$ such that*

$$Q(\mathbb{T}, \lambda) \;\leq\; c \max\{\lambda,\, A_1,\, A_2,\, \nu_3\, A_3\}. \tag{4.39}$$

The type of result achieved in the theorem is much like the result obtained in Bentkus, Bloznelis and Götze (1996). In fact, in this paper attention is restricted to the case where $c_j = 1$ for all $j$, and the truncation levels $M_j$ are all equal. Under these restrictions the Berry-Esseen bound reached for $\mathbb{T}$ is almost $\max\{A_1,\, A_2,\, A_3\}$ (a slightly better moment expression than $A_2$ is used). This shows that in the case where the $c_j$ equal 1, our concentration bound is quite acceptable. The restriction that the $M_j$ are equal is rather harsh, since it prevents one from applying the result to the two-sample Student statistic, and seems unnecessary for the proof. If so, then a nice Berry-Esseen bound for the two-sample Student statistic arises as well (cf. Corollary 4.13).

The remainder of the current subsection will constitute the proof of Theorem 4.5. To this, let

$$A \ := \ \max\{A_1,\, A_2,\, \nu_3\, A_3\}.$$

Taking the $c$ from (4.39), it is clear that the statement in the Theorem is trivially true in case $A \geq c^{-1}$, so we assume throughout that $A < c^{-1}$. We take $c \geq 1$. We start the proof by turning from our original statistic $\mathbb{T}$ to truncated versions of it. First we replace all $X_j$ by $U_j$. After this we get rid of the normalization by means of division by a square root, using a smooth function $g$ resembling the function $y \mapsto y^{-1/2}$. After this we are ready to apply Theorem 2.2.

Let

$$\tilde{\mathbb{T}} \ := \ \mathbb{T}(U) \ = \ (\textstyle\sum_{j=1}^{n} c_j\, U_j)/(\sum_{k=1}^{n} U_k^2 + R(U))^{1/2}. \tag{4.40}$$

Since, for any $x \in \mathbb{R}$ and $\lambda \geq 0$, using (4.33),

$$\begin{aligned} &\mathbb{P}(x \leq \mathbb{T} \leq x+\lambda) \\ &\quad\leq \ \mathbb{P}(x \leq \mathbb{T} \leq x+\lambda,\, X_j^2 \leq M_j^2,\ \text{all } j) + \mathbb{P}(X_j^2 > M_j^2,\ \text{some } j) \\ &\quad\leq \ \mathbb{P}(x \leq \tilde{\mathbb{T}} \leq x+\lambda) + A_1, \end{aligned} \tag{4.41}$$

it is clear that

$$Q(\mathbb{T},\lambda) \ \leq \ Q(\tilde{\mathbb{T}},\lambda) + A_1, \tag{4.42}$$

and we may turn our attention to finding $Q(\tilde{\mathbb{T}},\lambda)$.

Our second truncation takes place in order to get rid of the division by the square root. To this, let $g : [0,\infty) \to \mathbb{R}$ be a bounded, infinitely many times differentiable function with bounded derivatives, such that

$$g(y) \ = \ y^{-1/2} \qquad \text{for } y \in [\tfrac{1}{2},\infty). \tag{4.43}$$

Now let

$$V_k \ := \ U_k^2 - s_k^2, \tag{4.44}$$

for $k = 1, \dots, n$. Notice that $\mathbb{E} V_k = 0$ for all $k$ and

$$\operatorname{var} \textstyle\sum_{k=1}^n V_k \;=\; \sum_{k=1}^n \mathbb{E} V_k^2 \;\le\; \sum_{k=1}^n \mathbb{E} U_k^4 \;\le\; \sum_{k=1}^n \mathbb{E}|U_k|^3 \;=\; A_3. \tag{4.45}$$

Set

$$\tilde{\tilde{\mathbb{T}}} \;:=\; \big(\textstyle\sum_{j=1}^n c_j\, U_j\big)\, g\big(1 + \sum_{k=1}^n V_k + R(U)\big). \tag{4.46}$$

By Chebychev's inequality and (4.45),

$$\mathbb{P}\big(\big|\textstyle\sum_{k=1}^n V_k\big| \ge \frac{1}{4}\big) \;\le\; 16 \operatorname{var} \sum_{k=1}^n V_k \;\ll\; A_3, \tag{4.47}$$

whereas by assumption

$$\mathbb{P}(|R(U)| \ge \tfrac{1}{4}) \;\le\; 4\,\mathbb{E}|R(U)| \;\ll\; (\mathbb{E}|R(U)|^{3/2})^{2/3} \;\ll\; A_3^2.$$

As in (4.41), thus

$$\mathbb{P}(x \le \tilde{\mathbb{T}} \le x + \lambda) \;\ll\; \mathbb{P}(x \le \tilde{\tilde{\mathbb{T}}} \le x + \lambda) + A_3 \tag{4.48}$$

for all $x \in \mathbb{R}$, $\lambda \ge 0$, so that

$$Q(\tilde{\mathbb{T}}, \lambda) \;\ll\; Q(\tilde{\tilde{\mathbb{T}}}, \lambda) + A_3, \tag{4.49}$$

and it makes no difference for the theorem whether we prove (4.39) for $\tilde{\tilde{\mathbb{T}}}$ or $\tilde{\mathbb{T}}$.

For any $C \subset N$ we introduce the following two expressions:

$$\eta_{(C)} \;:=\; \textstyle\sum_{j\notin C} c_j\, U_j \quad\text{and}\quad \xi_{(C)} \;:=\; \sum_{j\notin C} V_j. \tag{4.50}$$

Moreover, let

$$\eta \;:=\; \textstyle\sum_{j=1}^n c_j\, U_j \quad\text{and}\quad \xi \;:=\; \sum_{j=1}^n V_j. \tag{4.51}$$

As in (4.45) we see that $\mathbb{E}\xi_{(C)}^2 \le A_3$ for any $C \subset N$. As concerns the moments of $\eta_{(C)}$ we have the following:

**Lemma 4.6.** *Assume that $s^2 \ne 0$, $M_j^2 \le s^2$ for all $j$ and $\sum_{j=1}^n c_j^2 = n$. Assume furthermore that either the sample is i.i.d. or $c_j = 1$ for all $j$, and that $A_2 \le 1$. Then, for any $C \subset N$,*

$$\mathbb{E}\big|\textstyle\sum_{j\in C} c_j\, U_j\big|^2 \;\ll\; 1, \quad \mathbb{E}\big|\sum_{j\in C} c_j\, U_j\big|^3 \;\ll\; \nu_3. \tag{4.52}$$

**Proof of Lemma 4.6.** Let $p \geq 2$ and let $Z_1, \dots, Z_n$ be a sequence of independent random variables with mean zero. Then Rosenthal's inequality (see Petrov (1995), Theorem 2.9) states that

$$\mathbb{E}\,|\sum_{j=1}^n Z_j|^p \;\ll_p\; \sum_{j=1}^n \mathbb{E}\,|Z_j|^p + (\sum_{j=1}^n \mathbb{E}\,Z_j^2)^{p/2}. \tag{4.53}$$

Now take any fixed $C \subset N$ and $p \in \{2, 3\}$. We take

$$Z_j \;:=\; c_j\,(U_j - \mathbb{E}\,U_j)\,I\{j \in C\},$$

for $1 \leq j \leq n$. Using the $C_r$-inequality (see, for example, Loève (1977), p.157) together with (4.53), we see that

$$\begin{aligned} \mathbb{E}\,|\sum_{j\in C} c_j\,U_j|^p &\ll \mathbb{E}\,|\sum_{j\in C} c_j\,(U_j - \mathbb{E}\,U_j)|^p + \mathbb{E}\,|\sum_{j\in C} c_j\,\mathbb{E}\,U_j|^p \\ &= \mathbb{E}\,|\sum_{j=1}^n Z_j|^p + |\sum_{j\in C} c_j\,\mathbb{E}\,U_j|^p \\ &\ll \sum_{j=1}^n \mathbb{E}\,|Z_j|^p + \big(\sum_{j=1}^n \mathbb{E}\,Z_j^2\big)^{p/2} + \big(\sum_{j\in C} |c_j|\,|\mathbb{E}\,U_j|\big)^p. \end{aligned} \tag{4.54}$$

Using Lyapunov's inequality and the fact that $|U_j| \leq 1$ we see that $|\mathbb{E}\,U_j|^p \leq \mathbb{E}\,|U_j|^p \leq \mathbb{E}\,U_j^2$ for all $j$, and a fortiori

$$\mathbb{E}\,|Z_j|^p \;\ll\; |c_j|^p\,(\mathbb{E}\,|U_j|^p + |\mathbb{E}\,U_j|^p) \;\ll\; |c_j|^p\,\mathbb{E}\,U_j^2.$$

Of course $\mathbb{E}\,Z_j^2 \leq c_j^2\,\mathbb{E}\,U_j^2$. In case $c_j = 1$ for all $j$, (4.54) leads to

$$\mathbb{E}\,|\sum_{j\in C} c_j\,U_j|^p \;\; \sum_{j=1}^n s_j^2 + \big(\sum_{j=1}^n s_j^2\big)^{p/2} + A_2^p \;\ll\; 1.$$

In the case that the sample is i.i.d. we have that $s_j^2 = n^{-1}$ and $|\mathbb{E}\,U_j| = n^{-1}A_2$, so that

$$\begin{aligned} \mathbb{E}\,|\sum_{j\in C} c_j\,U_j|^p \;\ll\; & n^{-1} \sum_{j=1}^n |c_j|^p + (n^{-1} \sum_{j=1}^n c_j^2)^{p/2} \\ & + (A_2\,n^{-1} \sum_{j=1}^n |c_j|)^p \;\ll\; \nu_p. \end{aligned}$$

This concludes the proof. □

Next we get rid of $R(U)$. To this, let

$$\hat{\mathbb{T}} \;:=\; (\sum_{j=1}^n c_j\,U_j)\,g(1 + \sum_{k=1}^n V_k) \;=\; \eta\,g(1+\xi). \tag{4.55}$$

Using Markov's and Hoeffding's inequality, (4.52) and the assumption on $\mathbb{E}\,|R(U)|^{3/2}$, we see that

$$\begin{aligned}\mathbb{P}(|\hat{\mathbb{T}}-\tilde{\tilde{\mathbb{T}}}|>A_3) &\leq A_3^{-1}\,\mathbb{E}\,|\hat{\mathbb{T}}-\tilde{\tilde{\mathbb{T}}}| \;\leq\; A_3^{-1}\,\|g'\|_\infty\,\mathbb{E}\,|\eta|\,|R(U)| \\ &\ll_g A_3^{-1}\,\{\mathbb{E}\,|\eta|^3\}^{1/3}\,\{\mathbb{E}\,|R(U)|^{3/2}\}^{2/3} \\ &\ll \nu_3\,A_3^{-1}\,A_3^2 \;=\; \nu_3\,A_3. \end{aligned} \tag{4.56}$$

As a result, as in (4.41),

$$\begin{aligned}\mathbb{P}(x\leq\tilde{\tilde{\mathbb{T}}}\leq x+\lambda) &\leq \mathbb{P}(x\leq\tilde{\tilde{\mathbb{T}}}\leq x+\lambda,\,|\hat{\mathbb{T}}-\tilde{\tilde{\mathbb{T}}}|\leq A_3)+\mathbb{P}(|\hat{\mathbb{T}}-\tilde{\tilde{\mathbb{T}}}|>A_3) \\ &\leq \mathbb{P}(x-A_3\leq\hat{\mathbb{T}}\leq x+A_3+\lambda)+\nu_3\,A_3, \end{aligned} \tag{4.57}$$

for any $x\in\mathbb{R}$ and $\lambda\geq 0$, and it is clear that

$$Q(\tilde{\tilde{\mathbb{T}}},\lambda) \;\leq\; Q(\hat{\mathbb{T}},\lambda+2A_3)+\nu_3\,A_3, \tag{4.58}$$

which reduces the problem to the estimation of $Q(\hat{\mathbb{T}},\lambda)$.

Now let $\hat{\sigma}^2 := \operatorname{var}\hat{\mathbb{T}}$. First we will show that $\hat{\sigma}^2$ is approximately 1, which will reduce the problem to the estimation of $Q(\hat{\mathbb{T}}/\hat{\sigma},\lambda)$. After that we derive bounds for $\beta(\hat{\mathbb{T}})$ and $\Delta^2(\hat{\mathbb{T}})$ and apply Theorem 2.2.

**Lemma 4.7.** *Let $f:\mathbb{R}\to\mathbb{R}$ be a two times differentiable function with bounded derivatives. Then, for any $x,y,z\in\mathbb{R}$, $|f(x+y)-f(x)|\leq\|f'\|_\infty\,|y|$ and*

$$|f(x+y+z)-f(x+y)-f(x+z)+f(x)| \;\leq\; \|f''\|_\infty\,|y|\,|z|.$$

*In general: if $f$ is $k$ times differentiable with a bounded $k^{th}$ derivative, and $x,y_1,\ldots,y_k\in\mathbb{R}$, then*

$$\Big|\textstyle\sum_{C\subset\{1,\ldots,k\}}(-1)^{|C|}\,f(x+\sum_{j\in\{1,\ldots,k\}\setminus C}y_j)\Big| \;\leq\; \|f^{(k)}\|_\infty\,|y_1|\cdots|y_k|.$$

**Proof of Lemma 4.7.** For any $1\leq l\leq k$ and $u\in\mathbb{R}$, we write

$$h(l,u) \;:=\; \textstyle\sum_{C\subset\{1,\ldots,l\}}(-1)^{|C|}\,f(u+\sum_{j\in\{1,\ldots,l\}\setminus C}y_j),$$

and we take $\tau_1,\ldots,\tau_k$ independently and uniformly distributed on (0,1). All the results are simple implications of the following equality:

$$h(l,u) \;=\; \mathbb{E}\,f^{(l)}(u+\textstyle\sum_{j=1}^{l}\tau_j\,y_j)\,y_1\cdots y_l. \tag{4.59}$$

The proof of (4.59) is given using induction. For $l = 1$ we have that

$$h(1,u) \;=\; f(u+y_1) - f(u) \;=\; \mathbb{E}\, f'(u+\tau_1\, y_1)\, y_1$$

following (2.8), and indeed (4.59) is correct. Now suppose that (4.59) applies for all $l \leq m-1$ and $u \in \mathbb{R}$. Distinguishing between the sets $C \subset \{1,\ldots,m\}$ that contain $m$ and the ones that do not, we see that

$$h(m,u) \;=\; h(m-1,u) - h(m-1,u+y_m).$$

As a consequence of the inductional hypothesis then

$$\begin{aligned} h(m,u) \;&=\; \mathbb{E}\{f^{(m-1)}(u + \textstyle\sum_{j=1}^{m-1} \tau_j\, y_j) \\ &\qquad - f^{(m-1)}(u + y_m + \textstyle\sum_{j=1}^{m-1} \tau_j\, y_j)\}\; y_1 \cdots y_{m-1} \end{aligned}$$

and, again using (2.8), indeed (4.59) for $l = m$. This concludes the proof. □

**Lemma 4.8.** *Assume that $A \leq 1$. There exist absolute constants $\tilde{c}_1$ and $\tilde{c}_2$ such that*

$$|\mathbb{E}\,\hat{\mathbb{T}}| \leq \tilde{c}_1\, A, \quad |\mathbb{E}\,\hat{\mathbb{T}}^2 - 1| \leq \tilde{c}_2\, A.$$

**Proof of Lemma 4.8.** First we take a look at $\mathbb{E}\,\hat{\mathbb{T}} = \sum_{j=1}^n c_j\, \mathbb{E}\, U_j\; g(1+\xi)$. Using Lyapunov's inequality, we have that in general

$$\begin{aligned} |\mathbb{E} U_j\, \{g(1+\xi) - g(1+\xi_{(j)})\}| \;&\leq\; \|g'\|_\infty\, \mathbb{E}\,|U_j|\,|V_j| \\ &\ll_g\; \mathbb{E}\,|U_j|^3 + s_j^2\, \mathbb{E}\,|U_j| \;\ll\; \mathbb{E}\,|U_j|^3, \end{aligned}$$

whereas $|\mathbb{E} U_j\; g(1+\xi_{(j)})| = |(\mathbb{E}\, U_j)\, \mathbb{E}\, g(1+\xi_{(j)})| \ll_g |\mathbb{E}\, U_j|$, so that

$$|\mathbb{E} U_j\; g(1+\xi)| \;\ll_g\; |\mathbb{E}\, U_j| + \mathbb{E}\,|U_j|^3.$$

In case $c_j = 1$ for all $j$ this leads to $|\mathbb{E}\,\hat{\mathbb{T}}| \ll_g A_2 + A_3$, which is what we are looking for. In the case that the sample is i.i.d. it follows that

$$|\mathbb{E}\,\hat{\mathbb{T}}| \;=\; |\mathbb{E} U_1\; g(1+\xi)|\,|\textstyle\sum_{j=1}^n c_j| \;\ll_g\; n\,\nu_1\,(|\mathbb{E}\, U_1| + \mathbb{E}\,|U_1|^3) \;\leq\; A_2 + A_3$$

as well, and this proves the first point.

For $\mathbb{E}\,\hat{\mathbb{T}}^2$ we first look at the case in which $c_j = 1$ for all $j$. In this case we have that

$$\mathbb{E}\,\hat{\mathbb{T}}^2 - 1 \;=\; \textstyle\sum_{j=1}^n \mathbb{E}\, U_j^2\; \{g^2(1+\xi) - 1\} + \sum_{j\neq l} \mathbb{E}\, U_j U_l\; g^2(1+\xi) \;=:\; d_1 + d_2,$$

with $|\mathbb{E}\,\hat{\mathbb{T}}^2 - 1| \le |d_1| + |d_2|$. As to $d_2$, writing $d_2^{jl} := \mathbb{E}\,U_j U_l\, g^2(1+\xi)$, we have the following:

$$\begin{aligned} |d_2^{jl}| \;\le\; & |\mathbb{E}\,U_j U_l \{g^2(1+\xi) - g^2(1+\xi_{(j)}) - g^2(1+\xi_{(l)}) + g^2(1+\xi_{(jl)})\}| \\ & + |\mathbb{E}\,U_j U_l \{g^2(1+\xi_{(j)}) - g^2(1+\xi_{(jl)})\}| \\ & + |\mathbb{E}\,U_j U_l \{g^2(1+\xi_{(l)}) - g^2(1+\xi_{(jl)})\}| \\ & + |\mathbb{E}\,U_j U_l\, g^2(1+\xi_{(jl)})| \;=:\; d_{21}^{jl} + d_{22}^{jl} + d_{23}^{jl} + d_{24}^{jl}. \end{aligned}$$

Clearly $d_{24}^{jl} = |\mathbb{E}\,U_j|\,|\mathbb{E}\,U_l|\,|\mathbb{E}\,g^2(1+\xi_{(jl)})| \ll_g |\mathbb{E}\,U_j|\,|\mathbb{E}\,U_l|$. Using Lemma 4.7,

$$d_{22}^{jl} \;\le\; \|(g^2)'\|_\infty\, |\mathbb{E}\,U_j|\, \mathbb{E}\,|U_l|\,|V_l| \;\ll_g\; |\mathbb{E}\,U_j|\, \mathbb{E}\,|U_l|^3,$$

similarly $d_{23}^{jl} \ll_g |\mathbb{E}\,U_l|\,\mathbb{E}\,|U_j|^3$, and moreover

$$d_{21}^{jl} \;\le\; \|(g^2)''\|_\infty\, \mathbb{E}\,|U_j U_l|\,|V_j V_l| \;\ll_g\; \mathbb{E}\,|U_j|^3\, \mathbb{E}\,|U_l|^3.$$

As a result

$$\begin{aligned} |d_2| \;\ll_g\; & \textstyle\sum_{j\ne l} \left(\mathbb{E}\,|U_j|^3\,\mathbb{E}\,|U_l|^3 + |\mathbb{E}\,U_j|\,\mathbb{E}\,|U_l|^3 + |\mathbb{E}\,U_l|\,\mathbb{E}\,|U_j|^3 + |\mathbb{E}\,U_j|\,|\mathbb{E}\,U_l|\right) \\ \le\; & \left(\textstyle\sum_{j=1}^n \mathbb{E}\,|U_j|^3\right)^2 + 2\left(\textstyle\sum_{j=1}^n |\mathbb{E}\,U_j|\right)\left(\textstyle\sum_{l=1}^n \mathbb{E}\,|U_l|^3\right) + \left(\textstyle\sum_{j=1}^n |\mathbb{E}\,U_j|\right)^2 \\ =\; & A_3^2 + 2A_2\,A_3 + A_2^2 \;\le\; 4A^2 \end{aligned}$$

and $|d_2| \ll_g A^2$. As to $d_1$, writing $d_1^j := \mathbb{E}\,U_j^2\left\{g^2(1+\xi) - 1\right\}$, we have that

$$\begin{aligned} |d_1^j| \;\le\; & \left|\mathbb{E}\,U_j^2\left\{g^2(1+\xi) - g^2(1+\xi_{(j)})\right\}\right| \\ & + (\mathbb{E}\,U_j^2)\left|\mathbb{E}\left(g^2(1+\xi_{(j)}) - g^2(1)\right)\right| \;=:\; d_{11}^j + d_{12}^j. \end{aligned}$$

Taking $\tau$ uniformly distributed on $(0,1)$ and independent of the sample, using (2.8) we see that

$$\begin{aligned} d_{12}^j \;=\; & s_j^2\left|\mathbb{E}\,(g^2)'(1)\,\xi_{(j)} + \mathbb{E}\,(g^2)''(1+\tau\xi_{(j)})\,\xi_{(j)}^2\right| \\ \le\; & s_j^2\,\|(g^2)''\|_\infty\,\mathbb{E}\,\xi_{(j)}^2 \;\ll_g\; s_j^2\,A_3, \end{aligned}$$

see the remark behind (4.51), whereas

$$d_{11}^j \;\le\; \|(g^2)'\|_\infty\,\mathbb{E}\,U_j^2\,|V_j| \;\ll_g\; \mathbb{E}\,|U_j|^3.$$

In turn

$$|d_1| \;\ll_g\; \textstyle\sum_{j=1}^n (\mathbb{E}\,|U_j|^3 + s_j^2\,A_3) \;=\; A_3 + A_3 \;\ll\; A_3,$$

and as a consequence $|\mathbb{E}\,\hat{\mathbb{T}}^2 - 1| \ll_g A$.

Finally we consider $\mathbb{E}\,\hat{\mathbb{T}}^2$ in the case of an i.i.d. sample. Going about in the same way,

$$\begin{aligned} \mathbb{E}\,\hat{\mathbb{T}}^2 - 1 \;&=\; \big(\textstyle\sum_{j=1}^n c_j^2\,\mathbb{E}\,U_j^2\; g^2(1+\xi)\big) - 1 + \sum_{j\neq l} c_j c_l\,\mathbb{E}\,U_j U_l\, g^2(1+\xi) \\ &=\; n\,\mathbb{E}\,U_1^2\,(g^2(1+\xi)-1) + \mathbb{E}\,U_1 U_2\, g^2(1+\xi)\,\textstyle\sum_{j\neq l} c_j c_l \;=:\; f_1 + f_2. \end{aligned}$$

As to $f_2$, using the analysis of $d_2^{jl}$ above,

$$|\mathbb{E}\,U_1 U_2\, g^2(1+\xi)| \;\ll_g\; \big(|\mathbb{E}\,U_1| + \mathbb{E}\,|U_1|^3\big)^2 \;\le\; n^{-2}\,(A_2+A_3)^2,$$

whereas $|\sum_{j\neq l} c_j c_l| \;\le\; (\sum_{j=1}^n |c_j|)^2 \;=\; n^2\,\nu_1^2 \;\le\; n^2$, and $|f_2| \ll_g A^2$. The analysis of $d_1^j$ above is showing us that

$$\mathbb{E}\,U_1^2\,(g^2(1+\xi)-1) \;\le\; \mathbb{E}\,|U_1|^3 + n^{-1}\,A_3 \;\ll\; n^{-1}\,A_3,$$

so that $|f_1| \ll_g A_3$, and as a consequence again $|\mathbb{E}\,\hat{\mathbb{T}}^2 - 1| \ll_g A$. This finishes the proof. □

Taking $\tilde{c}_1$ and $\tilde{c}_2$ as in Lemma 4.8, we take the constant $c$ such that

$$1 - \tilde{c}_2\,c^{-1} - \tilde{c}_1^2\,c^{-2} \;\ge\; \tfrac{1}{2}.$$

In this way

$$\hat{\sigma}^2 \;=\; 1 + (\mathbb{E}\,\hat{\mathbb{T}}^2 - 1) - (\mathbb{E}\,\hat{\mathbb{T}})^2 \;\ge\; 1 - \tilde{c}_2\,A - \tilde{c}_1^2\,A^2 \;\ge\; \tfrac{1}{2}. \tag{4.60}$$

Using Theorem 2.2 we see that

$$\begin{aligned} Q(\hat{\mathbb{T}},\lambda) \;=\; Q(\hat{\mathbb{T}}/\hat{\sigma},\lambda/\hat{\sigma}) \;&\ll\; \max\{\lambda\,\hat{\sigma}^{-1},\, \beta(\hat{\mathbb{T}})\,\hat{\sigma}^{-3},\, \Delta^2(\hat{\mathbb{T}})\,\hat{\sigma}^{-2}\} \\ &\ll\; \max\{\lambda,\, \beta(\hat{\mathbb{T}}),\, \Delta^2(\hat{\mathbb{T}})\}, \end{aligned} \tag{4.61}$$

and we turn to finding bounds for $\beta(\hat{\mathbb{T}})$ and $\Delta^2(\hat{\mathbb{T}})$.

**Lemma 4.9.** *Assume that $A_3 \le c^{-1}$. We have:*

$$\beta(\hat{\mathbb{T}}) \ll_g \nu_3\,A_3, \quad \Delta^2(\hat{\mathbb{T}}) \ll_g A_3. \tag{4.62}$$

**Proof of Lemma 4.9.** Let $W := (W_1, \dots, W_n)$ be an independent copy of $U$

(that is to say, $U$ and $W$ are independent and $U \stackrel{d}{=} W$), and let $Z_j := W_j^2 - s_j^2$, for $1 \le j \le n$.

First we look at $\beta(\hat{\mathbb{T}})$. Notice that with (A.13) we have that

$$\beta(\hat{\mathbb{T}}) \;\le\; \textstyle\sum_{j=1}^n \mathbb{E}\,|D_j\hat{\mathbb{T}}|^3. \tag{4.63}$$

Now take any $1 \le j \le n$ and write $S_1 := c_j\, U_j\, g(1+\xi)$, $S_2 := \eta_{(j)}\, g(1+\xi)$. It is clear that $\hat{\mathbb{T}} = S_1 + S_2$, and using (A.9) and (A.10) we see that

$$D_j\hat{\mathbb{T}} \;=\; D_jS_1 + D_jS_2 \;=\; c_j\, D_jU_j\, g(1+V_j+\xi_{(j)}) + \eta_{(j)}\, D_jg(1+V_j+\xi_{(j)}),$$

so that with the $C_r$-inequality

$$|D_j\hat{\mathbb{T}}|^3 \;\ll\; |c_j|^3\, |D_jU_j\, g(1+V_j+\xi_{(j)})|^3 + |\eta_{(j)}|^3\, |D_jg(1+V_j+\xi_{(j)})|^3. \tag{4.64}$$

As to the first term on the right side of (4.64),

$$\begin{aligned}
|D_jU_j\, g(1+V_j+\xi_{(j)})| \;&=\; |\mathbb{E}\,(U_j\, g(1+V_j+\xi_{(j)}) - W_j\, g(1+Z_j+\xi_{(j)})\,|\,U)| \\
&\le\; |\mathbb{E}\,(U_j\,\{g(1+V_j+\xi_{(j)}) - g(1+Z_j+\xi_{(j)})\}\,|\,U)| \\
&\qquad\qquad + |\mathbb{E}\,(g(1+Z_j+\xi_{(j)})\,\{U_j - W_j\}\,|\,U)| \\
&\le\; |U_j|\,\|g'\|_\infty\,\mathbb{E}\,(|V_j - Z_j|\,|\,V_j) + \|g\|_\infty\,\mathbb{E}\,(|U_j - W_j|\,|\,U_j)| \\
&\ll_g\; |U_j|\,(|V_j| + \mathbb{E}\,|V_j|) + |U_j| + \mathbb{E}\,|U_j| \;\ll\; |U_j| + s_j.
\end{aligned}$$

As to the second one,

$$\begin{aligned}
|D_jg(1+V_j+\xi_{(j)})| \;&=\; |\mathbb{E}\,(g(1+V_j+\xi_{(j)}) - g(1+Z_j+\xi_{(j)})\,|\,U)| \\
&\le\; \|g'\|_\infty\,\mathbb{E}\,(|V_j - Z_j|\,|\,U_j) \\
&\ll_g\; U_j^2 + s_j^2 \;\ll\; |U_j| + s_j.
\end{aligned} \tag{4.65}$$

As a result

$$|D_j\hat{\mathbb{T}}|^3 \;\ll_g\; (|c_j|^3 + |\eta_{(j)}|^3)\,(|U_j|^3 + s_j^3), \tag{4.66}$$

and, using in turn (4.63), Lemma 4.6 and Lyapunov's inequality,

$$\beta(\hat{\mathbb{T}}) \;\ll_g\; \textstyle\sum_{j=1}^n (|c_j|^3 + \nu_3)\,(\mathbb{E}\,|U_j|^3 + s_j^3).$$

In case all $c_j = 1$ this means that

$$\beta(\hat{\mathbb{T}}) \;\ll_g\; \textstyle\sum_{j=1}^n \mathbb{E}\,|U_j|^3 = A_3 = \nu_3\, A_3.$$

In the case of an i.i.d. sample, $\beta(\hat{\mathbb{T}}) \ll_g (\mathbb{E}|U_1|^3 + n^{-3/2})\, n\,\nu_3 \;\ll\; \nu_3\, A_3$ as well, which proves (4.62) for $\beta(\hat{\mathbb{T}})$.

We turn to

$$\Delta^2(\hat{\mathbb{T}}) \;=\; \textstyle\sum_{1\le j<k\le n} \mathbb{E}\,|D_{jk}\hat{\mathbb{T}}|^2.$$

We take any combination $1 \le j < k \le n$ and write

$$S_1 := c_j\, U_j\, g(1+\xi), \quad S_2 := c_k\, U_k\, g(1+\xi) \quad \text{and} \quad S_3 := \eta_{(jk)}\, g(1+\xi),$$

so that

$$\begin{aligned} D_{jk}\hat{\mathbb{T}} &= D_{jk}S_1 + D_{jk}S_2 + D_{jk}S_3 \\ &= \textstyle\sum_{l=j,k} c_l\, D_{jk}U_l\, g(1+\xi) + \eta_{(jk)}\, D_{jk} g(1+\xi). \end{aligned} \tag{4.67}$$

Taking $l = j$, we see that

$$\begin{aligned} &|D_{jk}U_j\, g(1+\xi)| \;=\; |D_jU_j\, D_k g(1+\xi)| \\ &\quad = \; |\mathbb{E}\,(U_j\, D_k g(1+V_j+\xi_{(j)}) - W_j\, D_k g(1+Z_j+\xi_{(j)}) \,|\, U)| \\ &\quad \le \; |U_j|\, |D_k g(1+V_j+V_k+\xi_{(jk)})| \\ &\qquad\qquad + \mathbb{E}\,(|W_j|\, |D_k g(1+Z_j+V_k+\xi_{(jk)})| \,|\, U), \end{aligned} \tag{4.68}$$

and since, as in (4.65), $|D_k g(1+X+V_k+\xi_{(jk)})| \ll_g U_k^2 + s_k^2$ for $X = V_j, Z_j$, we see that

$$|D_{jk}S_1| \;\ll_g\; |c_j|\, (|U_j| + s_j)\, (U_k^2 + s_k^2). \tag{4.69}$$

Of course we have the similar bound $|D_{jk}S_2| \ll_g |c_k|\,(|U_k| + s_k)\,(U_j^2 + s_j^2)$. Using Lemma 4.7 we see that

$$\begin{aligned} &|D_{jk} g(1+\xi)| \\ &\quad = \; |\mathbb{E}\,(g(1+V_j+V_k+\xi_{(jk)}) - g(1+V_j+Z_k+\xi_{(jk)}) \\ &\qquad\qquad - g(1+Z_j+V_k+\xi_{(jk)}) + g(1+Z_j+Z_k+\xi_{(jk)}) \;|\; U)| \\ &\quad \le \; \|g''\|_\infty\, \mathbb{E}\,(|V_j - Z_j|\, |V_k - Z_k| \,|\, U) \;\ll_g\; (U_j^2 + s_j^2)\,(U_k^2 + s_k^2). \end{aligned} \tag{4.70}$$

In conclusion,

$$\begin{aligned} \Delta^2(\hat{\mathbb{T}}) \;&\ll_g\; \textstyle\sum_{1\le j<k\le n} \left(c_j^2\, s_j^2\, \mathbb{E}\,U_k^4 + c_k^2\, s_k^2\, \mathbb{E}\,U_j^4 + \mathbb{E}\,|\eta_{(jk)}|^2\, \mathbb{E}\,U_j^4\, \mathbb{E}\,U_k^4\right) \\ &\ll\; \left(\textstyle\sum_{j=1}^n c_j^2\, s_j^2\right) \left(\textstyle\sum_{k=1}^n \mathbb{E}\,U_k^4\right) + \left(\textstyle\sum_{k=1}^n \mathbb{E}\,U_k^4\right)^2 \;\ll\; A_3, \end{aligned} \tag{4.71}$$

using the fact that in both of our cases $\sum_{j=1}^n c_j^2\, s_j^2 = 1$ and that $|U_k| \le 1$. Thus (4.62) for $\Delta^2(\hat{\mathbb{T}})$, and this finishes the proof. □

Using in turn (4.42), (4.49), (4.58), (4.61) and Lemma 4.9 we see that indeed (4.39), which finishes the proof of Theorem 4.5.

## 4.5 Self-normalized statistics: Berry-Esseen bounds

We turn to finding a Berry-Esseen bound for the self-normalized statistic $\mathbb{T}$, as defined in (4.34). We go about as in the previous section, now starting from Theorem 3.3. As the result of this theorem is relatively worse than that of Theorem 2.2 because of the $\varepsilon_1$ and $\varepsilon_2$ in the powers of $\Delta_1^2\, s^{-2}$ and $\Delta_2^2\, s^{-2}$, we are not able to obtain the Berry-Esseen bound $\max\{A_1, A_2, \nu_3\, A_3\}$ over here, which would have been a direct generalization of Theorem 4.5, giving us the best results known. Instead, we will assume the existence of the moment $\mathbb{E}\,|X_j|^{3+\delta}$ for all $j$, for some $\delta > 0$, which will lead us to bounds that are rather close to the best results.

We formulate the result. We look at $\mathbb{T}$ as in (4.34), with $R = R(X)$ satisfying (4.33) and $c_1, \dots, c_n$ satisfying $\sum_{j=1}^n c_j^2 = n$, starting either from an i.i.d. sample or from an independent sample such that $c_j = 1$ for all $j$. We use the notation $s^2$, $U_j$, $s_j^2$ as in (4.35) and (4.36), $A_1$, $A_2$ and $A_3$ as in (4.37), and, for some fixed $0 < \delta < 1$, define the moment expression

$$A_4 \ := \ \textstyle\sum_{j=1}^n \mathbb{E}\,|U_j|^{3+\delta}.$$

Now we have the following:

**Theorem 4.10.** *Assume that $s^2 \neq 0$, $M_j \leq s$ for all $j$ and $\sum_{j=1}^n c_j^2 = n$. Moreover, assume that either the sample $X_1, \dots, X_n$ is i.i.d. or that $c_j = 1$ for all $j$. Assume that (4.33) applies and $\mathbb{E}\,|R(U)|^{3/2} \ll A_3^3$, and let $\delta \in (0,1)$ be fixed. Then there exists a constant $c(\delta)$ such that*

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T} \leq x) - \Phi(x)| \ \leq \ c(\delta)\ \max\{A_1,\, A_2,\, \nu_3\, A_3,\, A_4^{1/(1+\delta)}\}. \tag{4.72}$$

For a discussion of the result, see the previous section, following the statement of Theorem 4.5. The $A_4^{1/(1+\delta)}$ is worse than the bound achieved in Bentkus, Bloznelis and Götze (1996), because instead of $3^{\mathrm{rd}}$ moments now $(3+\delta)^{\mathrm{th}}$ moments are used.

**Proof of Theorem 4.10.** We go about as in Section 4.4, now using Theorem 3.3 instead of Theorem 2.2.

First we set

$$A \ := \ \max\left\{A_1,\, A_2,\, \nu_3\, A_3,\, A_4^{1/(1+\delta)}\right\}.$$

As before we may assume that $A \leq c^{-1}$ and that $c$ is such that $|\hat{\sigma}^2 - 1| \leq \frac{1}{2}$, cf. (4.60). Notice that, as $|U_j| \leq 1$ for all $j$, $A_4 \leq A_3$. We start by making the same transitions as in Section 4.4.

First we go from $\mathbb{T}$ to $\tilde{\mathbb{T}} = \mathbb{T}(U)$, cf. (4.40). To this, note that, taking $Z$ standard normally distributed and independent of the sample, for any $x \in \mathbb{R}$ we have that

$$\begin{aligned}
|\mathbb{P}(\mathbb{T} \leq x) - \Phi(x)| \;&=\; |\mathbb{E}\,(I\{\mathbb{T} \leq x\} - I\{Z \leq x\})| \\
\leq\;& |\mathbb{E}\,I\{X_j^2 \leq M_j^2,\ \text{all } j\}\,(I\{\mathbb{T} \leq x\} - I\{Z \leq x\})| \\
&\quad + \mathbb{E}\,I\{X_j^2 > M_j^2,\ \text{some } j\} \\
\leq\;& A_1 + |\mathbb{E}\,(1 - I\{X_j^2 > M_j^2,\ \text{some } j\})\,(I\{\tilde{\mathbb{T}} \leq x\} - I\{Z \leq x\})| \\
\leq\;& 2A_1 + |\mathbb{P}(\tilde{\mathbb{T}} \leq x) - \Phi(x)|, \qquad (4.73)
\end{aligned}$$

see (4.41). This implies that for our purposes it makes no difference whether we look at $\mathbb{T}$ or $\tilde{\mathbb{T}}$.

Our second transition is from $\tilde{\mathbb{T}}$ to $\tilde{\tilde{\mathbb{T}}}$, as defined in (4.46), using $g$ and $V_k$ as before. Indeed, with (4.47) and the remark following it

$$\mathbb{P}\big(\big|\textstyle\sum_{k=1}^n V_k + R(U)\big| \geq \frac{1}{2}\big) \;\leq\; \mathbb{P}\big(\big|\sum_{k=1}^n V_k\big| \geq \frac{1}{4}\big) + \mathbb{P}\big(|R(U)| \geq \frac{1}{4}\big) \;\ll\; A_3,$$

and as in (4.73) we see that

$$|\mathbb{P}(\tilde{\mathbb{T}} \leq x) - \Phi(x)| \;\ll\; |\mathbb{P}(\tilde{\tilde{\mathbb{T}}} \leq x) - \Phi(x)| + A_3. \qquad (4.74)$$

Next we remove $R(U)$ from $\tilde{\tilde{\mathbb{T}}}$ and get to $\hat{\mathbb{T}}$, as defined in (4.55). In fact, using (4.56) we see that

$$\begin{aligned}
\mathbb{P}(\tilde{\tilde{\mathbb{T}}} \leq x) - \Phi(x) \;&=\; \mathbb{E}\,(I\{\tilde{\tilde{\mathbb{T}}} \leq x\} - I\{Z \leq x\}) \\
\leq\;& \mathbb{P}(|\tilde{\tilde{\mathbb{T}}} - \hat{\mathbb{T}}| > A_3) + \mathbb{E}\,I\{|\tilde{\tilde{\mathbb{T}}} - \hat{\mathbb{T}}| \leq A_3\}\,(I\{\hat{\mathbb{T}} \leq x + A_3\} - I\{Z \leq x\}) \\
\ll\;& \nu_3\,A_3 + \big(\mathbb{P}(\hat{\mathbb{T}} \leq x + A_3) - \Phi(x)\big) + \mathbb{P}(|\tilde{\tilde{\mathbb{T}}} - \hat{\mathbb{T}}| > A_3) \\
\ll\;& \nu_3\,A_3 + |\mathbb{P}(\hat{\mathbb{T}} \leq x + A_3) - \Phi(x + A_3)| + A_3 \\
\ll\;& \nu_3\,A_3 + \sup\nolimits_{y \in \mathbb{R}} |\mathbb{P}(\hat{\mathbb{T}} \leq y) - \Phi(y)|, \qquad (4.75)
\end{aligned}$$

whereas in the same way

$$\begin{aligned}
\Phi(x) - \mathbb{P}(\tilde{\tilde{\mathbb{T}}} \leq x) \;&\ll\; \nu_3\,A_3 + (\Phi(x) - \mathbb{P}(\hat{\mathbb{T}} \leq x - A_3)) \\
&\ll\; \nu_3\,A_3 + \sup\nolimits_{y \in \mathbb{R}} |\mathbb{P}(\hat{\mathbb{T}} \leq y) - \Phi(y)|, \qquad (4.76)
\end{aligned}$$

so that indeed

$$|\mathbb{P}(\tilde{\hat{\mathbb{T}}} \le x) - \Phi(x)| \ \ll\ \nu_3\, A_3 + \sup_{x\in\mathbb{R}} |\mathbb{P}(\hat{\mathbb{T}} \le x) - \Phi(x)|, \tag{4.77}$$

and instead of $\tilde{\hat{\mathbb{T}}}$ we may look at $\hat{\mathbb{T}}$.

As promised, to $\hat{\mathbb{T}}$ we apply Theorem 3.5, taking

$$\varepsilon_1 \ = \ \varepsilon_2 \ = \ \delta/(1+\delta).$$

Here we have to derive upper bounds for the moment expressions $\beta(\hat{\mathbb{T}})$, $\Delta_1^2(\hat{\mathbb{T}})$, $\Delta_2^2(\hat{\mathbb{T}})$ and

$$\tilde{\kappa} \ := \ \textstyle\sum_{1\le j<k\le n} \mathbb{E}\, T_j T_k\, T_{j,k},$$

after which we will have the required result for $(\hat{\mathbb{T}} - \mathbb{E}\,\hat{\mathbb{T}})/\hat{\sigma}$. The proof is then easily concluded. We turn to the moment expressions.

First note that following (4.62) and (4.71) it follows that

$$\beta(\hat{\mathbb{T}}) \ \ll_g \ \nu_3\, A_3 \quad \text{and} \quad \Delta_1^2(\hat{\mathbb{T}}) \le \ \Delta^2(\hat{\mathbb{T}}) \ \ll_g \ A_4. \tag{4.78}$$

As to $\Delta_2^2(\hat{\mathbb{T}})$ the bound $A_4$ is not sharp enough, and we will turn to third instead of second differences in order to obtain the right order.

In fact we have that

$$\tfrac{1}{3}\Delta_2^2(\hat{\mathbb{T}}) \ \le \ \textstyle\sum_{1\le j<k<l\le n} \mathbb{E}\, |D_{jkl}\,\hat{\mathbb{T}}|^2 \ =: \ \Delta_3^2(\hat{\mathbb{T}}). \tag{4.79}$$

Indeed, from (A.12) it follows that

$$\Delta_3^2(\hat{\mathbb{T}}) \ = \ \textstyle\sum_{A:|A|=3} \sum_{C:A\subset C} \mathbb{E}\, T_C^2 \ = \ \sum_{l=3}^n \binom{l}{3} \sum_{C:|C|=l} \mathbb{E}\, T_C^2,$$

whereas $\frac{1}{3}\binom{l}{2} \le \binom{l}{3}$ for $3 \le l \le n$, and using (A.19), (4.79) follows easily. As a result we may concentrate on $\Delta_3^2(\hat{\mathbb{T}})$. Here we go about as in the proof of Lemma 4.9. Let, for $p = j, k, l$,

$$S_p := c_p U_p\, g(1+\xi), \quad \text{and} \quad S_{jkl} := \eta_{(jkl)}\, g(1+\xi).$$

As in (4.68) here

$$\begin{aligned} |D_{jkl} S_j| \ \ &\le \ \ |c_j|\, |U_j|\, |D_{kl} g(1 + V_j + \xi_{(j)})| \\ &\qquad + |c_j|\, \mathbb{E}\, (|W_j|\, |D_{kl} g(1 + Z_j + \xi_{(j)})|\, |\, U), \end{aligned}$$

where, as in (4.70),

$$|D_{kl}g(1+X+\xi_{(j)})| \ \ll_g \ (U_k^2+s_k^2)\,(U_l^2+s_l^2)$$

for $X = V_j, Z_j$, so that

$$|D_{jkl}S_j| \ \ll_g \ |c_j|\,(U_k^2+s_k^2)\,(U_l^2+s_l^2)\,(|U_j|+s_j). \tag{4.80}$$

Analogue bounds apply for $S_k$ and $S_l$. As in (4.70), now using Lemma 4.7 for $k=3$, we see that

$$\begin{aligned} |D_{jkl}S_{jkl}| \ &= \ |\eta_{(jkl)}|\,|D_{jkl}g(1+\xi)| \\ &\ll_g \ |\eta_{(jkl)}|\,(U_j^2+s_j^2)\,(U_k^2+s_k^2)\,(U_l^2+s_l^2). \end{aligned} \tag{4.81}$$

Collecting the estimates, using (4.80), (4.81), Lemma 4.6 and symmetry we see that

$$\begin{aligned} \Delta_3^2(\hat{\mathbb{T}}) \ &\ll_g \ \textstyle\sum_{1\le j,k,l\le n}(c_j^2\,s_j^2\,\mathbb{E}U_k^4\,\mathbb{E}U_l^4+\mathbb{E}U_j^4\,\mathbb{E}U_k^4\,\mathbb{E}U_l^4) \\ &\le \ (\textstyle\sum_{j=1}^n c_j^2\,s_j^2)\,(\sum_{k=1}^n \mathbb{E}U_k^4)^2+(\sum_{j=1}^n \mathbb{E}U_j^4)^3 \ \ll \ A_4^2, \end{aligned}$$

cf. (4.71), and it follows that

$$\Delta_2^2(\hat{\mathbb{T}}) \ \ll_g \ A_4^2. \tag{4.82}$$

Next we look at $\tilde\kappa$. Using (A.11), in the special case of $\hat{\mathbb{T}}$ we may obtain a better bound than (3.10). In fact, using (4.66) and Lemma 4.6 we see that

$$\begin{aligned} |T_p| \ &= \ |\mathbb{E}\,(D_p\hat{\mathbb{T}}\,|\,p)| \ \le \ \mathbb{E}\,(|D_p\hat{\mathbb{T}}|\,|\,p) \\ &\ll_g \ \mathbb{E}\,((|c_p|+|\eta_{(p)}|)\,(|U_p|+s_p)\,|\,p) \ \ll \ (|c_p|+1)\,(|U_p|+s_p) \end{aligned}$$

for $p = j,k$, and, following (4.67), (4.69) and (4.70), that

$$\begin{aligned} |T_{jk}| \ &\le \ \mathbb{E}\,(|D_{jk}\hat{\mathbb{T}}|\,|\,j,k) \\ &\ll_g \ \mathbb{E}\,(|c_j|\,(|U_j|+s_j)\,(U_k^2+s_k^2)+|c_k|\,(|U_k|+s_k)\,(U_j^2+s_j^2)\,|\,j,k) \\ &\qquad + \mathbb{E}\,(|\eta_{(jk)}|\,(U_j^2+s_j^2)\,(U_k^2+s_k^2)\,|\,j,k) \\ &\ll \ |c_j|\,(|U_j|+s_j)\,(U_k^2+s_k^2)+|c_k|\,(|U_k|+s_k)\,(U_j^2+s_j^2) \\ &\qquad + (U_j^2+s_j^2)\,(U_k^2+s_k^2). \end{aligned}$$

It is easily seen from this that

$$\begin{aligned}
\tilde{\kappa} &\ll_g \big(\textstyle\sum_{j=1}^n |c_j|\,(|c_j|+1)\,\mathbb{E}\,(|U_j|+s_j)^2\big) \\
&\qquad\qquad \big(\textstyle\sum_{k=1}^n (|c_k|+1)\,\mathbb{E}\,(U_k^2+s_k^2)\,(|U_k|+s_k)\big) \\
&\qquad\qquad + \big(\textstyle\sum_{j=1}^n (|c_j|+1)\,\mathbb{E}\,(|U_j|+s_j)\,(U_j^2+s_j^2)\big)^2 \\
&\ll \big(\textstyle\sum_{j=1}^n (c_j^2+|c_j|)\,s_j^2\big)\,\big(\textstyle\sum_{k=1}^n (|c_k|+1)\,\mathbb{E}\,|U_k|^3\big) \\
&\qquad\qquad + \big(\textstyle\sum_{j=1}^n (|c_j|+1)\,\mathbb{E}\,|U_j|^3\big)^2.
\end{aligned}$$

As a consequence, in case $c_j = 1$ for all $j$, $\tilde{\kappa} \ll_g A_3 + A_3^2 \ll A_3$, whereas in the case of an i.i.d. sample as well

$$\begin{aligned}
\tilde{\kappa} &\ll_g n^{-1}\,(n + \textstyle\sum_{j=1}^n |c_j|)^2\,\mathbb{E}\,|U_1|^3 + (\mathbb{E}\,|U_1|^3)^2\,(n+\textstyle\sum_{j=1}^n |c_j|)^2 \\
&\ll \big(n\,\mathbb{E}\,|U_1|^3 + (n\,\mathbb{E}\,|U_1|^3)^2\big)\,(1+n^{-1}\textstyle\sum_{j=1}^n |c_j|)^2 \ll A_3 + A_3^2 \ll A_3,
\end{aligned}$$

using that $n^{-1}\sum_{j=1}^n |c_j| \le (n^{-1}\sum_{j=1}^n c_j^2)^{1/2} = 1$, cf. (4.38). As a result,

$$\tilde{\kappa} \ll_g A_3. \tag{4.83}$$

Following (4.78), (4.82), (4.83) and the fact that by construction $\hat{\sigma}^2 \ge \frac{1}{2}$, and taking $\varepsilon_1$ and $\varepsilon_2$ as proposed, now from (3.9) we conclude that

$$\begin{aligned}
D(\hat{\mathbb{T}}) &\ll_\delta \tilde{\kappa}\,\hat{\sigma}^{-3}\,\|\Phi'''\|_\infty \\
&\qquad + \max\big\{\beta(\hat{\mathbb{T}})\,\hat{\sigma}^{-3},\ (\Delta_1^2(\hat{\mathbb{T}})\,\hat{\sigma}^{-2})^{1/(1+\delta)},\ (\Delta_2^2(\hat{\mathbb{T}})\,\hat{\sigma}^{-2})^{1/2(1+\delta)}\big\} \\
&\ll_g \nu_3\,A_3 + \max\{\nu_3\,A_3,\ A_4^{1/(1+\delta)}\} \ll A,
\end{aligned} \tag{4.84}$$

and we only need to show that for our purposes $(\hat{\mathbb{T}} - \mathbb{E}\,\hat{\mathbb{T}})/\hat{\sigma}$ is approximately equal to $\hat{\mathbb{T}}$. To this, take any $x \in \mathbb{R}$. Using Lemma 4.8 and (4.84), we see that, taking $y := (x - \mathbb{E}\,\hat{\mathbb{T}})/\hat{\sigma}$,

$$\begin{aligned}
|\mathbb{P}(\hat{\mathbb{T}} \le x) - \Phi(x)| &\le |\mathbb{P}((\hat{\mathbb{T}} - \mathbb{E}\,\hat{\mathbb{T}})/\hat{\sigma} \le y) - \Phi(y)| + |\Phi(y) - \Phi(x)| \\
&\le D(\hat{\mathbb{T}}) + |\Phi(y) - \Phi(x/\hat{\sigma})| + |\Phi(x/\hat{\sigma}) - \Phi(x)| \\
&\ll D(\hat{\mathbb{T}}) + |\mathbb{E}\,\hat{\mathbb{T}}|/\hat{\sigma} + |\Phi(x/\hat{\sigma}) - \Phi(x)| \\
&\ll_{g,\delta} A + |\Phi(x/\hat{\sigma}) - \Phi(x)|,
\end{aligned} \tag{4.85}$$

and the only thing left to do is to find a uniform bound for $|\Phi(x/\hat{\sigma}) - \Phi(x)|$. Here with (2.8), taking $\tau$ uniformly distributed on $(0,1)$, we have that

$$\begin{aligned}
|\Phi(x/\hat{\sigma}) - \Phi(x)| &= |\mathbb{E}\,\Phi'(x + \tau\,(\hat{\sigma}^{-1}-1)\,x)\,(\hat{\sigma}^{-1}-1)\,x| \\
&\le \hat{\sigma}^{-1}\,|1-\hat{\sigma}|\,(1+\hat{\sigma})\,f(x) \ll f(x)\,A,
\end{aligned}$$

using Lemma 4.8 and the lower bound for $\hat\sigma^2$, and writing

$$f(x) \;:=\; \int_0^1 |x|\,\exp\{-\tfrac12 x^2\,(1+t\,(\hat\sigma^{-1}-1))^2\}\,dt.$$

Here, for all $0 \le t \le 1$,

$$|1+t\,(\hat\sigma^{-1}-1)| \;\ge\; 1-t\,|\hat\sigma^{-1}-1| \;\ge\; 1-(\sqrt{2}-1) \;\ge\; \tfrac12,$$

since $\frac12 \le \hat\sigma^2 \le \frac32$, so that $f(x) \le |x|\,\exp\{-\frac18 x^2\} \ll 1$, and indeed

$$|\Phi(x/\hat\sigma)-\Phi(x)| \;\ll\; A. \tag{4.86}$$

Now (4.72) is a simple consequence of (4.73), (4.74), (4.77), (4.85) and (4.86), and this finishes the proof. □

## 4.6 Student's statistic

Our first application of the Theorems 4.5 and 4.10 will be on Student's statistic, in a one-sample and a two-sample setting.

Here and in the next subsection we will need the following lemma, which, together with its proof, can be found as Lemma 1.3 in Bentkus and Götze (1996a).

**Lemma 4.11.** *Let $\alpha$ be any real-valued random variable, and $k$ any integer. There exists a largest solution $q_0^2 = q_0^2(\alpha,k)$ of the following equation in $q^2$:*

$$q^2 \;=\; \mathbb{E}\,\alpha^2\,I\{\alpha^2 \le q^2\,k\}. \tag{4.87}$$

*We have that $q_0^2 \le \mathbb{E}\,\alpha^2$, which may equal $\infty$, and in case $q_0^2 \le \frac12\,\mathbb{E}\,\alpha^2$ we have that $\mathbb{E}\,\alpha^2\,I\{\alpha^2 \le \frac12\,\mathbb{E}\,\alpha^2\,k\} \le \frac12\,\mathbb{E}\,\alpha^2$.*

We first look at the classical Student statistic. To this, let $X_1,\dots,X_n$ be an independent, identically distributed sample, with $\mathbb{E}\,X_1 = 0$ and $0 < \sigma^2 := \mathbb{E}\,X_1^2 < \infty$. We look at self-normalized sums

$$\mathbb{T} \;:=\; \frac{X_1+\ldots+X_n}{\left(X_1^2+\ldots+X_n^2\right)^{1/2}},$$

or, alternatively, at Student's statistic

$$\mathbb{T}_0 \;:=\; \frac{X_1+\ldots+X_n}{\left\{\sum_{k=1}^n (X_k-\overline{X})^2\right\}^{1/2}} \;=\; \frac{X_1+\ldots+X_n}{\left(\sum_{k=1}^n X_k^2 - n\,\overline{X}^2\right)^{1/2}},$$

with $\overline{X}$ denoting the sample mean. In the case of a normal distribution, that is, if the $X_j$ are $N(0,\sigma^2)$-distributed, the statistic $(1-n^{-1})^{1/2}\,\mathbb{T}_0$ is $t$-distributed with $n-1$ degrees of freedom, that is to say, almost $N(0,1)$-distributed.

Now let $q^2$ be the $q_0^2(\alpha,k)$ which we obtain from Lemma 4.11 when taking $\alpha = X_1$ and $k=n$, let

$$B \;:=\; n^{-1/2}\,q^{-3}\,\mathbb{E}\,|X_1|^3\,I\{X_1^2 \le q^2\,n\} + n^{1/2}\,q^{-1}\,\mathbb{E}\,|X_1|\,I\{X_1^2 > q^2\,n\},$$

and, for any fixed constant $\delta \in (0,1)$, let

$$\begin{aligned} C \;&:=\; n^{-1/2}\left(q^{-(3+\delta)}\,\mathbb{E}\,|X_1|^{3+\delta}\,I\{X_1^2 \le q^2\,n\}\right)^{3/(3+\delta)} \\ &\qquad\qquad + n^{1/2}\,q^{-1}\,\mathbb{E}\,|X_1|\,I\{X_1^2 > q^2\,n\}. \end{aligned}$$

We have the following result:

**Corollary 4.12.** *Suppose that the sample $X_1,\dots,X_n$ is i.i.d., $\mathbb{E}\,X_1 = 0$ and $q^2 \neq 0$. Let $\delta \in (0,1)$ be some fixed constant. Then there exist constants $c_1$ and $c_2(\delta)$ such that, for Student's statistic $\mathbb{T}$, we have the concentration bound*

$$Q(\mathbb{T},\lambda) \;\le\; c_1\,\max\{\lambda, B\}, \tag{4.88}$$

*and the Berry-Esseen bound*

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T} \le x) - \Phi(x)| \;\le\; c_2(\delta)\,C. \tag{4.89}$$

*In case that the sample is i.i.d., $\mathbb{E}\,X_1 = 0$ and $0 < \sigma^2 < \infty$, this leads us to the concentration bound*

$$Q(\mathbb{T},\lambda) \;\le\; c_1\,\max\{\lambda\,,\,n^{-1/2}\,\sigma^{-3}\,\mathbb{E}\,|X_1|^3\}, \tag{4.90}$$

*and the Berry-Esseen bound*

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T} \le x) - \Phi(x)| \;\le\; c_2(\delta)\,n^{-1/2}\,\sigma^{-(3+\delta)}\,\mathbb{E}\,|X_1|^{3+\delta}. \tag{4.91}$$

*The same results hold for $\mathbb{T}_0$.*

Notice that the bounds achieved in (4.88) and (4.89) in comparison with the bounds (4.90) and (4.91) have the nice property of being already finite if $\mathbb{E}\,|X_1|$ is finite. The best known results in this case are the Berry-Esseen bounds (4.89) and (4.91) for $\delta = 0$, see Theorem 1.2 in Bentkus and Götze (1996a), or Corollary 1.2 in Bentkus, Bloznelis and Götze (1996). This indicates that (4.88) and (4.90) are the proper concentration bounds. The Berry-Esseen bounds (4.89) and (4.91) are worse than the ones mentioned, but they seem to be the best results achieved using a theory on general symmetric asymptotically normal statistics, the best result so far being by Friedrich (1989), whose (3.7) leads to (4.89) for $\delta = \frac{1}{3}$.

Next we look at the so-called two-sample Student statistic. To this, let $Y_1, \ldots, Y_{n_1}$ and $Z_1, \ldots, Z_{n_2}$ be two independent, i.i.d. samples of real-valued variables, for which

$$\mu_1 := \mathbb{E}\, Y_1, \quad \mu_2 := \mathbb{E}\, Z_1, \quad \sigma_1^2 := \operatorname{var} Y_1 \quad \text{and} \quad \sigma_2^2 := \operatorname{var} Z_1.$$

For any sample $W_1, \ldots, W_m$, let

$$\overline{W} := m^{-1} \textstyle\sum_{j=1}^m W_j \quad \text{and} \quad S_W^2 := m^{-1} \textstyle\sum_{j=1}^m (W_j - \overline{W})^2.$$

In cases where one is interested whether the two populations have the same mean, the two-sample Student statistic

$$S \;:=\; \frac{\overline{Y} - \overline{Z}}{\sqrt{n_1^{-1} S_Y^2 + n_2^{-1} S_Z^2}}$$

is a natural test statistic. Writing

$$V_1 \;:=\; Y_1 - \mu_1, \qquad V_2 \;:=\; Z_1 - \mu_2,$$

we take $q_1^2$ and $q_2^2$ as the $q_0^2(\alpha, k)$ which we obtain from Lemma 4.11 when taking $(\alpha, k) = (V_1, n_1)$ and $(\alpha, k) = (V_2, n_2)$ respectively. Now define $B_k = B_k(n_k)$ by

$$B_k \;:=\; n_k^{-1/2}\, q_k^{-3}\, \mathbb{E}\, |V_k|^3\, I\{V_k^2 \le q_k^2\, n_k\} + n_k^{1/2}\, q_k^{-1}\, \mathbb{E}\, |V_k|\, I\{V_k^2 > q_k^2\, n_k\},$$

and, for any fixed constant $\delta \in (0,1)$, define $C_k = C_k(n_k)$ by

$$\begin{aligned} C_k \;:=\;& n_k^{-1/2} \left(q_k^{-(3+\delta)}\, \mathbb{E}\, |V_k|^{3+\delta}\, I\{V_k^2 \le q_k^2\, n_k\}\right)^{3/(3+\delta)} \\ &\qquad + n_k^{1/2}\, q_k^{-1}\, \mathbb{E}\, |V_k|\, I\{V_k^2 > q_k^2\, n_k\}, \end{aligned}$$

for $k = 1, 2$. We have the following result:

**Corollary 4.13.** *Assume that $\mu_1 = \mu_2$ and that $q_1^2, q_2^2 \neq 0$. Let $\delta \in (0,1)$ be some fixed constant. Then for the two-sample Student statistic $S$ there exist constants $c_1$ and $c_2(\delta)$ such that*

$$Q(S, \lambda) \;\leq\; c_1 \max\{\lambda,\, B_1,\, B_2\}, \tag{4.92}$$

*and*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(S \leq x) - \Phi(x)| \;\leq\; c_2(\delta) \max\{C_1,\, C_2\}. \tag{4.93}$$

*In the case that $0 < \sigma_1^2, \sigma_2^2 < \infty$, taking $\mu := \mu_1 = \mu_2$, this means that constants $c_1$ and $c_2(\delta)$ exist such that we have the concentration bound*

$$Q(S, \lambda) \;\leq\; c_1 \max\{\lambda\,,\, n_1^{-1/2}\, \sigma_1^{-3}\, \mathbb{E}\,|V_1|^3\,,\, n_2^{-1/2}\, \sigma_2^{-3}\, \mathbb{E}\,|V_2|^3\}, \tag{4.94}$$

*and the Berry-Esseen bound*

$$\begin{aligned} &\sup_{x \in \mathbb{R}} |\mathbb{P}(S \leq x) - \Phi(x)| \\ &\quad\leq\; c_2(\delta) \max\{n_1^{-1/2}\, \sigma_1^{-(3+\delta)}\, \mathbb{E}\,|V_1|^{3+\delta},\, n_2^{-1/2}\, \sigma_2^{-(3+\delta)}\, \mathbb{E}\,|V_2|^{3+\delta}\}. \end{aligned} \tag{4.95}$$

No Berry-Esseen bounds seem to have been established for $S$ yet, so Corollary 4.13 would be the best result available. The concentration bounds look nice since they are very similar to the ones obtained in the case of one sample. The Berry-Esseen bounds seem to be good apart from the $\delta$. As mentioned earlier, it seems to be no problem to adjust the results of Bentkus, Bloznelis and Götze (1996) in such a way that they become applicable as well for $S$, after which it is easy to derive (4.93) and (4.95) for $\delta = 0$.

We turn to the proofs of the Corollaries 4.12 and 4.13.

**Proof of Corollary 4.12.** We want to apply Theorem 4.5 and 4.10, in the case that $c_j = 1$ for all $j$. Note that under this condition $\mathbb{T}$ and $\mathbb{T}_0$ are of the form (4.34), with $R(X) = 0$ and $R(X) = -n\,\overline{X}^2$ respectively. We define $M_j^2 := q^2\, n$, for $1 \leq j \leq n$. In this way

$$s^2 \;=\; n\,\mathbb{E}\,X_1^2\, I\{X_1^2 \leq q^2\, n\} \;=\; n\, q^2 \tag{4.96}$$

(see Lemma 4.11), so that indeed $M_j^2 \le s^2$ for all $j$. Note that as well by assumption $s^2 = n\,q^2 > 0$, so Theorem 4.5 is indeed applicable. As to $A_1$:

$$A_1 \;=\; n\,\mathbb{P}(X_1^2 \ge q^2\,n) \;\le\; n^{1/2}\,q^{-1}\,\mathbb{E}\,|X_1|\,I\{X_1^2 > q^2\,n\} \;\le\; B. \tag{4.97}$$

Since $\mathbb{E}\,X_1 = 0$, we have that $\mathbb{E}\,X_1\,I\{X_1^2 \le q^2\,n\} = -\mathbb{E}\,X_1\,I\{X_1^2 > q^2\,n\}$, so that

$$\begin{aligned} A_2 \;&=\; n\,s^{-1}\,|\mathbb{E}\,X_1\,I\{X_1^2 > q^2\,n\}| \\ &\le\; n^{1/2}\,q^{-1}\,\mathbb{E}\,|X_1|\,I\{X_1^2 > q^2\,n\} \;\le\; B, \end{aligned} \tag{4.98}$$

and in the same way

$$A_3 \;\le\; n^{-1/2}\,q^{-3}\,\mathbb{E}\,|X_1|^3\,I\{X_1^2 \le q^2\,n\} \;\le\; B.$$

Finally note that in the case that $R(X) = n\,\overline{X}^2$, using Lemma 4.6 and (4.38),

$$\mathbb{E}\,|R(U)|^{3/2} \;=\; n^{-3/2}\,\mathbb{E}\,|\eta|^3 \;\ll\; n^{-3/2} \;\le\; A_3^3$$

as required, and the concentration bound (4.88) is an easy consequence.

Now let, for $s \ge 0$,

$$\theta_s \;:=\; q^{-s}\,\mathbb{E}\,|X_1|^s\,I\{X_1^2 \le q^2\,n\}.$$

With Lyapunov's inequality we see that

$$1 = \theta_2 \;\le\; \theta_3^{2/3} \;\le\; \theta_{3+\delta}^{2/(3+\delta)}. \tag{4.99}$$

Now

$$\begin{aligned} A_4 \;=\; n\,\mathbb{E}\,|U_1|^{3+\delta} \;&=\; n\,(n\,q^2)^{-(3+\delta)/2}\,\mathbb{E}\,|X_1|^{3+\delta}\,I\{X_1^2 \le q^2\,n\} \\ &=\; n^{-(1+\delta)/2}\,\theta_{3+\delta}, \end{aligned}$$

and using (4.99) it follows that

$$A_4^{1/(1+\delta)} \;=\; n^{-1/2}\,\theta_{3+\delta}^{1/(1+\delta)} \;\le\; n^{-1/2}\,\theta_{3+\delta}^{3/(3+\delta)} \;\le\; C.$$

It follows as well from (4.99) that $B \le C$, and (4.89) is an easy consequence of Theorem 4.10.

We turn to the proof of (4.90) and (4.91). Here without loss of generality we may assume that $q^2 > \frac{1}{2}\sigma^2$. Indeed, in the opposite case Lemma 4.11 is telling us that $\mathbb{E}X_1^2\, I\{X_1^2 > \frac{1}{2}\sigma^2 n\} \geq \frac{1}{2}\sigma^2$, so that

$$\begin{aligned} 1 \;&\leq\; (\tfrac{1}{2}\sigma^2)^{-1}\, \mathbb{E}X_1^2\, I\{X_1^2 > \tfrac{1}{2}\sigma^2 n\} \qquad (4.100)\\ &\leq\; n^{-1/2}\, (\tfrac{1}{2}\sigma^2)^{-3/2}\, \mathbb{E}|X_1|^3\, I\{X_1^2 > \tfrac{1}{2}\sigma^2 n\} \;\leq\; 2^{3/2}\, n^{-1/2}\, \sigma^{-3}\, \mathbb{E}|X_1|^3, \end{aligned}$$

and a fortiori, again using Lyapunov's inequality,

$$1 \;\leq\; 2^{3/2}\, n^{-1/2} \left(\sigma^{-(3+\delta)}\, \mathbb{E}|X_1|^{3+\delta}\right)^{3/(3+\delta)} \;\leq\; 2^{3/2}\, n^{-1/2}\, \sigma^{-(3+\delta)}\, \mathbb{E}|X_1|^{3+\delta},$$

which would make the two statements trivially true.

Under the assumption that $q^2 > \frac{1}{2}\sigma^2$,

$$\begin{aligned} B \;&\leq\; n^{-1/2}\, q^{-3}\, \mathbb{E}|X_1|^3\, I\{X_1^2 \leq q^2 n\}\\ &\qquad + n^{1/2}\, q^{-1}\, (q^2 n)^{-1}\, \mathbb{E}|X_1|^3\, I\{X_1^2 > q^2 n\}\\ &=\; n^{-1/2}\, q^{-3}\, \mathbb{E}|X_1|^3 \;\leq\; 2^{3/2}\, n^{-1/2}\, \sigma^{-3}\, \mathbb{E}|X_1|^3 \end{aligned}$$

and

$$\begin{aligned} C \;&\leq\; n^{-1/2}\, \theta_{3+\delta}^{3/(3+\delta)} + n^{1/2}\, q^{-1}\, (q^2 n)^{-(2+\delta)/2}\, \mathbb{E}|X_1|^{3+\delta}\, I\{X_1^2 > q^2 n\}\\ &\leq\; n^{-1/2}\, \theta_{3+\delta} + n^{-(1+\delta)/2}\, q^{-(3+\delta)}\, \mathbb{E}|X_1|^{3+\delta}\, I\{X_1^2 > q^2 n\}\\ &\leq\; n^{-1/2}\, q^{-(3+\delta)}\, \mathbb{E}|X_1|^{3+\delta} \;\leq\; 4\, n^{-1/2}\, \sigma^{-(3+\delta)}\, \mathbb{E}|X_1|^{3+\delta}, \end{aligned}$$

and (4.90) and (4.91) are obvious consequences of (4.88) and (4.89). This finishes the proof of the corollary. □

**Proof of Corollary 4.13.** Without loss of generality we will suppose that $\mu = 0$. Indeed, if not we may apply the reduced theorem to the situation in which all $Y_j$ and $Z_j$ are replaced by $Y_j - \mu$ and $Z_j - \mu$, and obtain the desired result.

Let

$$X_j \;:=\; n_1^{-1}\, Y_j \quad \text{and} \quad X_{n_1+k} \;:=\; -n_2^{-1}\, Z_k,$$

for $1 \leq j \leq n_1$ and $1 \leq k \leq n_2$. Unfortunately we are not able to apply Theorem 4.5 and 4.10 directly to our statistic $S$, as the appropriate $R(X)$ is not small enough, but the theorems *are* in fact applicable to the very similar statistic

$$\mathbb{T} \;=\; \frac{X_1 + \ldots + X_{n_1+n_2}}{\left(X_1^2 + \ldots + X_{n_1+n_2}^2\right)^{1/2}} \;=\; \frac{\overline{Y} - \overline{Z}}{\left(n_1^{-2}\, \sum_{k=1}^{n_1} Y_k^2 + n_2^{-2}\, \sum_{k=1}^{n_2} Z_k^2\right)^{1/2}}.$$

In order to apply the theorems, we first define

$$M_j^2 \;:=\; q_1^2\,n_1^{-1} \quad \text{and} \quad M_{n_1+k}^2 \;:=\; q_2^2\,n_2^{-1}$$

for $1 \le j \le n_1$ and $1 \le k \le n_2$. We write $U_j$, $s_j^2$ and so on as earlier. By construction

$$\begin{aligned} s^2 &= n_1\,n_1^{-2}\,\mathbb{E}\,Y_1^2\,I\{Y_1^2 \le q_1^2\,n_1\} + n_2\,n_2^{-2}\,\mathbb{E}\,Z_1^2\,I\{Z_1^2 \le q_2^2\,n_2\} \\ &= n_1^{-1}\,q_1^2 + n_2^{-1}\,q_2^2, \end{aligned}$$

and clearly $M_j^2 \le s^2$ for all $j$. As a consequence of the assumption that $q_1^2, q_2^2 \neq 0$ we have that $s^2 \neq 0$. From here on we denote

$$B \;:=\; \max\{B_1, B_2\} \quad \text{and} \quad C \;:=\; \max\{C_1, C_2\}.$$

We first take a look at $A_1$, $A_2$, $A_3$ and $A_4$. As in (4.97) we have that

$$A_1 \;=\; n_1\,\mathbb{P}(Y_1^2 \ge q_1^2\,n_1) + n_2\,\mathbb{P}(Z_1^2 \ge q_2^2\,n_2) \;\le\; B_1 + B_2 \;\ll\; B.$$

As in (4.98),

$$\begin{aligned} A_2 &= n_1\,|\mathbb{E}\,U_1| + n_2\,|\mathbb{E}\,U_{n_1+1}| \\ &= s^{-1}\,|\mathbb{E}\,Y_1\,I\{Y_1^2 > q_1^2\,n_1\}| + s^{-1}\,|\mathbb{E}\,Z_1\,I\{Z_1^2 > q_2^2\,n_2\}| \;\ll\; B, \end{aligned}$$

and in the same way $A_3 \ll B$. Note that, as in the proof of Corollary 4.12,

$$B_k \le C_k \quad \text{and} \quad B \le C.$$

Now let, for $k = 1, 2$ and $s \ge 0$,

$$\theta_{k,s} \;:=\; q_k^{-s}\,\mathbb{E}\,|V_k|^s\,I\{V_k^2 \le q_k^2\,n_k\}.$$

As in (4.99), for $k = 1, 2$, we have that $1 \le \theta_{k,3} \le \theta_{k,3+\delta}^{3/(3+\delta)}$. As to

$$A_4 \;=\; n_1\,\mathbb{E}\,|U_1|^{3+\delta} + n_2\,\mathbb{E}\,|U_{n_1+1}|^{3+\delta}$$

we notice that

$$\begin{aligned} \mathbb{E}\,|U_1|^{3+\delta} &= (n_1^{-1}\,q_1^2 + n_2^{-1}\,q_2^2)^{-(3+\delta)/2}\,\mathbb{E}\,|X_1|^{3+\delta}\,I\{X_1^2 \le q_1^2\,n_1^{-1}\} \\ &\le n_1^{(3+\delta)/2}\,q_1^{-(3+\delta)}\,n_1^{-(3+\delta)}\,\mathbb{E}\,|Y_1|^{3+\delta}\,I\{Y_1^2 \le q_1^2\,n_1\} \\ &= n_1^{-(3+\delta)/2}\,\theta_{1,3+\delta}, \end{aligned}$$

and in the same way $\mathbb{E}\,|U_{n_1+1}|^{3+\delta} \le n_2^{-(3+\delta)/2}\,\theta_{2,3+\delta}$, so that

$$A_4 \;\le\; n_1^{-(1+\delta)/2}\,\theta_{1,3+\delta} + n_2^{-(1+\delta)/2}\,\theta_{2,3+\delta},$$

and

$$\begin{aligned} A_4^{1/(1+\delta)} &\;\ll\; n_1^{-1/2}\,\theta_{1,3+\delta}^{1/(1+\delta)} + n_2^{-1/2}\,\theta_{2,3+\delta}^{1/(1+\delta)} \\ &\;\le\; n_1^{-1/2}\,\theta_{1,3+\delta}^{3/(3+\delta)} + n_2^{-1/2}\,\theta_{2,3+\delta}^{3/(3+\delta)} \;\le\; 2\,C. \end{aligned}$$

As a consequence

$$\max\{A_1,\, A_2,\, \nu_3\, A_3\} \;\ll\; B \quad \text{and} \quad \max\{A_1,\, A_2,\, \nu_3\, A_3,\, A_4^{1/(1+\delta)}\} \;\ll\; C, \tag{4.101}$$

which are the right bounds for our purposes.

Now take

$$\overline{X}_{(1)} := n_1^{-1}\,\textstyle\sum_{j=1}^{n_1} X_j, \quad \overline{X}_{(2)} := n_2^{-1}\,\textstyle\sum_{j=n_1+1}^{n_1+n_2} X_j,$$

and as well $\overline{U}_{(1)} := n_1^{-1}\,\sum_{j=1}^{n_1} U_j$ and $\overline{U}_{(2)} := n_2^{-1}\,\sum_{j=n_1+1}^{n_1+n_2} U_j$. We have that

$$\begin{aligned} S &= \big(\textstyle\sum_{j=1}^{n_1+n_2} X_j\big)\big/\big(\textstyle\sum_{k=1}^{n_1}(X_k-\overline{X}_{(1)})^2 + \textstyle\sum_{k=n_1+1}^{n_1+n_2}(X_k-\overline{X}_{(2)})^2\big)^{1/2} \\ &= \big(\textstyle\sum_{j=1}^{n_1+n_2} X_j\big)\big/\big(\textstyle\sum_{k=1}^{n_1+n_2} X_k^2 + R(X)\big)^{1/2}, \end{aligned}$$

that is to say, $S$ is of the form (4.34), taking $c_j = 1$ and

$$R(X) \;:=\; -(n_1\,\overline{X}_{(1)}^2 + n_2\,\overline{X}_{(2)}^2).$$

We set

$$\tilde{S} \;:=\; S(U),$$

and, taking $g$ and $V_k$ as in (4.43) and (4.44),

$$\overline{S} \;:=\; \big(\textstyle\sum_{j=1}^{n_1+n_2} U_j\big)\, g(1 + \textstyle\sum_{k=1}^{n_1+n_2} V_k + R(U)).$$

First we will prove that

$$Q(S,\lambda) \;\ll\; Q(\overline{S},\lambda) + B \tag{4.102}$$

and

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(S \le x) - \Phi(x)| \ \ll\ \sup_{x\in\mathbb{R}} |\mathbb{P}(\overline{S} \le x) - \Phi(x)| + C. \qquad (4.103)$$

Indeed, (4.42) is telling us that

$$Q(S,\lambda) \ \le\ Q(\tilde{S},\lambda) + A_1 \ \ll\ Q(\tilde{S},\lambda) + B,$$

whereas by (4.73)

$$|\mathbb{P}(S \le x) - \Phi(x)| \ \le\ 2A_1 + |\mathbb{P}(\tilde{S} \le x) - \Phi(x)| \ \ll\ |\mathbb{P}(\tilde{S} \le x) - \Phi(x)| + C,$$

so instead of $S$ we may look at $\tilde{S}$. To get from $\tilde{S}$ to $\overline{S}$ we need the following. Because

$$\mathbb{E}\, U_1^2 \ =\ s^{-2}\, n_1^{-2}\, \mathbb{E}\, Y_1^2\, I\{Y_1^2 \le q_1^2\, n_1\} \ =\ s^{-2}\, n_1^{-2}\, q_1^2,$$

by Lyapunov's inequality $\mathbb{E}\,|U_1|^3 \ge s^{-3}\, n_1^{-3}\, q_1^3$, and as a consequence

$$n_1^{-1/2} \ \le\ s^3\, q_1^{-3}\, n_1^{5/2}\, \mathbb{E}\,|U_1|^3 \ =\ q_1^{-3}\, n_1^{-1/2}\, \mathbb{E}\,|Y_1|^3\, I\{Y_1^2 \le q_1^2\, n_1\} \ \le\ 2^{3/2}\, B_1. \qquad (4.104)$$

Of course as well $n_2^{-1/2} \le 2^{3/2}\, B_2$, and using Markov's inequality together with Lemma 4.6,

$$\begin{aligned} \mathbb{P}(|R(U)| \ge \tfrac{1}{4}) \quad &\le\quad 4\,(n_1^{-1}\, \mathbb{E}\,|\textstyle\sum_{j=1}^{n_1} U_j|^2 + n_2^{-1}\, \mathbb{E}\,|\textstyle\sum_{j=n_1+1}^{n_1+n_2} U_j|^2) \\ &\ll\quad n_1^{-1} + n_2^{-1} \ \ll\ B_1^2 + B_2^2 \ \ll\ B^2. \end{aligned}$$

Moreover, by (4.47), $\mathbb{P}(|\sum_{k=1}^{n_1+n_2} V_k| \ge \frac{1}{4}) \le 16A_3 \ll B$, and as in (4.48) and (4.49) we may conclude that

$$Q(\tilde{S},\lambda) \ \ll\ Q(\overline{S},\lambda) + B,$$

which proves that indeed (4.102). As in (4.74)

$$|\mathbb{P}(\tilde{S} \le x) - \Phi(x)| \ \ll\ |\mathbb{P}(\overline{S} \le x) - \Phi(x)| + B \ \le\ |\mathbb{P}(\overline{S} \le x) - \Phi(x)| + C,$$

which proves (4.103).

Now let

$$\hat{S} \ :=\ \big(\textstyle\sum_{j=1}^{n_1+n_2} U_j\big)\, g(1 + \textstyle\sum_{k=1}^{n_1+n_2} V_k).$$

From the proofs of Theorems 4.5 and 4.10 it is clear that

$$Q(\hat{S},\lambda) \;\ll\; \max\{\lambda,\, A_2,\, A_3\} \;\ll\; \max\{\lambda, B\}$$

and, see as well (4.101),

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\hat{S} \le x) - \Phi(x)| \;\ll_{g,\delta}\; \max\{A_2,\, \nu_3\, A_3,\, A_4^{1/(1+\delta)}\} \;\ll\; C.$$

Taking $\eta$ as in (4.51), $\eta_{(1)} := \sum_{j=1}^{n_1} U_j$ and $\eta_{(2)} := \sum_{j=n_1+1}^{n_1+n_2} U_j$, using Markov's inequality and the foregoing we see that

$$\begin{aligned}
\mathbb{P}(|\hat{S} - \overline{S}| > B) \quad &\le \quad B^{-1}\,\mathbb{E}\,|\hat{S} - \overline{S}| \;\le\; B^{-1}\,\|g'\|_\infty\,\mathbb{E}\,|\eta|\,(n_1^{-1}\,\eta_{(1)}^2 + n_2^{-1}\,\eta_{(2)}^2) \\
&\ll_g \quad B^{-1}\,\big(n_1^{-1}\,\mathbb{E}\,|\eta_{(1)}|^3 + n_1^{-1}\,\mathbb{E}\,|\eta_{(2)}|\,\mathbb{E}\,\eta_{(1)}^2 \\
&\qquad\qquad + n_2^{-1}\,\mathbb{E}\,|\eta_{(1)}|\,\mathbb{E}\,\eta_{(2)}^2 + n_2^{-1}\,\mathbb{E}\,|\eta_{(2)}|^3\big) \\
&\ll \quad (n_1^{-1} + n_2^{-1})\,B^{-1} \;\ll\; (B_1^2 + B_2^2)\,B^{-1} \;\ll\; B,
\end{aligned}$$

and, a fortiori, $\mathbb{P}(|\hat{S} - \overline{S}| > C) \ll C$. As in (4.57) and (4.58) this leads to the fact that

$$Q(\overline{S},\lambda) \;\ll\; Q(\hat{S},\lambda + 2B) + B \;\ll\; \max\{\lambda, B\},$$

which concludes the proof of (4.92). On the other hand, as in (4.75) and (4.76) we see that

$$|\mathbb{P}(\overline{S} \le x) - \Phi(x)| \;\ll\; C + \sup_{x\in\mathbb{R}} |\mathbb{P}(\hat{S} \le x) - \Phi(x)| \;\ll_{g,\delta}\; C,$$

which concludes the proof of (4.93).

Finally we turn to the proof of (4.94) and (4.95). Here we may assume that

$$q_k^2 \;\ge\; \tfrac{1}{2}\sigma_k^2$$

for $k = 1, 2$. Indeed, if this is not the case then as in (4.100) we have that

$$1 \;\le\; 2^{3/2}\, n_k^{-1/2}\, \sigma_k^{-3}\, \mathbb{E}\,|V_k|^3 \;\le\; 2^{3/2}\, n_k^{-1/2}\, \sigma_k^{-(3+\delta)}\, \mathbb{E}\,|V_k|^{3+\delta}$$

for some $k \in \{1, 2\}$, and the statements will be trivially true. As in the proof of Corollary 4.12 this easily leads to (4.94) and (4.95). $\square$

## 4.7 Linear regression

In the theory of linear regression we look at so-called response variables $Y_1, \ldots, Y_n$ which are of the form

$$Y_j \;:=\; \alpha + \beta\, x_j + E_j \quad (1 \le j \le n),$$

where the $\alpha$, $\beta$, $x_1, \ldots, x_n$ are fixed real numbers, and $E_1, \ldots, E_n$ is a sequence of independent, identically distributed variables with

$$\mathbb{E}\, E_1 = 0 \quad \text{and} \quad \sigma^2 \;:=\; \mathbb{E}\, E_1^2 \;<\; \infty.$$

The numbers $\alpha$ and $\beta$ are supposed to be unknown, the sequence $x_1, \ldots, x_n$ is known, and we are looking for estimators for $\alpha$ and $\beta$.

We assume that not $x_1 = \ldots = x_n$, write $\overline{x}$, $\overline{Y}$ and $\overline{E}$ as usual and, for $p = 1, 2, 3$, set

$$\rho_p \;:=\; n^{-1} \textstyle\sum_{j=1}^n |x_j - \overline{x}|^p.$$

With Lyapunov's inequality $\rho_1^2 \le \rho_2$ and $\rho_2^{3/2} \le \rho_3$.

The following two estimators:

$$\hat\alpha \;:=\; \overline{Y} - \hat\beta\, \overline{x} \quad \text{and} \quad \hat\beta \;:=\; (n\rho_2)^{-1} \textstyle\sum_{j=1}^n (x_j - \overline{x})\, Y_j,$$

present the least square estimators for $\alpha$ and $\beta$, in the sense that they minimize the sum of squared errors

$$\mathrm{SSE} \;:=\; \textstyle\sum_{j=1}^n (Y_j - (\hat\alpha + \hat\beta\, x_j))^2.$$

For classical results, see e.g. Pestman (1998), Chapter 4. We have for example that

$$\mathbb{E}\,\hat\alpha = \alpha, \quad \mathbb{E}\,\hat\beta = \beta, \quad \operatorname{var}\hat\alpha = (n\rho_2)^{-1}\,(\rho_2 + \overline{x}^2)\,\sigma^2, \quad \operatorname{var}\hat\beta = (n\rho_2)^{-1}\,\sigma^2,$$

and we look at the statistics

$$\theta_1 \;:=\; \left(\frac{n\rho_2}{\rho_2 + \overline{x}^2}\right)^{1/2} \frac{\hat\alpha - \alpha}{\sqrt{\mathrm{SSE}/n}} \quad \text{and} \quad \theta_2 \;:=\; (n\rho_2)^{1/2}\, \frac{\hat\beta - \beta}{\sqrt{\mathrm{SSE}/n}}.$$

In the case of normal regression, that is to say, if the $E_j$ are $N(0, \sigma^2)$-distributed, the statistics $(1 - 2n^{-1})^{1/2}\, \theta_k$ are $t$-distributed with $n - 2$ degrees of freedom.

Let from here on

$$B \;:=\; n^{-1/2}\, q^{-3}\, \mathbb{E}\,|E_1|^3\, I\{E_1^2 \le q^2\, n\} + q^{-1}\, \mathbb{E}\,|E_1|\, I\{E_1^2 > q^2\, n\}$$

and, for any fixed constant $\delta \in (0,1)$,

$$\begin{aligned} C \;&:=\; n^{-1/2}\, \big(q^{-(3+\delta)}\, \mathbb{E}\,|E_1|^{3+\delta}\, I\{E_1^2 \le q^2\, n\}\big)^{3/(3+\delta)} \\ &\qquad\qquad + n^{1/2}\, q^{-1}\, \mathbb{E}\,|E_1|\, I\{E_1^2 > q^2\, n\}. \end{aligned}$$

We have the following result:

**Corollary 4.14.** *Assume that $0 < \sigma^2 < \infty$ and let $\delta \in (0,1)$ be some fixed constant. Then, for $k = 1, 2$, there exist constants $c_1$ and $c_2(\delta)$ such that*

$$Q(\theta_k, \lambda) \;\le\; c_1 \, \max\big\{\lambda\,,\, \rho_3\, \rho_2^{-3/2}\, B\big\} \tag{4.105}$$

*and*

$$\sup\nolimits_{x\in\mathbb{R}} |\mathbb{P}(\theta_k \le x) - \Phi(x)| \;\le\; c_2(\delta)\, \rho_3\, \rho_2^{-3/2}\, C. \tag{4.106}$$

*As a result, for $k = 1, 2$,*

$$Q(\theta_k, \lambda) \;\le\; c\, \max\big\{\lambda\,,\, \rho_3\, \rho_2^{-3/2}\, n^{-1/2}\, \sigma^{-3}\, \mathbb{E}\,|E_1|^3\big\} \tag{4.107}$$

*and*

$$\sup\nolimits_{x\in\mathbb{R}} |\mathbb{P}(\theta_k \le x) - \Phi(x)| \;\le\; c_2(\delta)\, \rho_3\, \rho_2^{-3/2}\, n^{-1/2}\, \sigma^{-(3+\delta)}\, \mathbb{E}\,|E_1|^{3+\delta}. \tag{4.108}$$

The best known results in this case are the Berry-Esseen bounds (4.106) and (4.108) for $\delta = 0$, see Bloznelis and Rackauskas (1999), Theorems 2.1 and 1.1. Hence, our concentration bounds are acceptable, whereas again our Berry-Esseen bounds are ‘almost’ right.

Corollary 4.14 will now be proved as an application of the Theorems 4.5 and 4.10. Without loss of generality we assume that $B, C \le 1$. First we need to rewrite $\theta_1$ and $\theta_2$ in terms of $E := (E_1, \ldots, E_n)$. Let from here on, for $j = 1, \ldots, n$,

$$v_j \;:=\; \rho_2^{1/2}\, (\rho_2 + \overline{x}^2)^{-1/2}\, (1 - \rho_2^{-1}\, \overline{x}\, (x_j - \overline{x})) \quad \text{and} \quad w_j \;:=\; \rho_2^{-1/2}\, (x_j - \overline{x}).$$

**Lemma 4.15.** *We have the following equalities:*

$$\begin{cases} \hat\beta - \beta &= n^{-1}\,\rho_2^{-1/2}\,\sum_{j=1}^n w_j E_j, \\ \hat\alpha - \alpha &= \overline{E} - (\hat\beta-\beta)\,\overline{x} \;=\; n^{-1}\,\rho_2^{-1/2}\,(\rho_2+\overline{x}^2)^{1/2}\,\sum_{j=1}^n v_j E_j, \\ SSE &= \sum_{j=1}^n E_j^2 - n\,\overline{E}^2 - n^{-1}\,(\sum_{j=1}^n w_j E_j)^2. \end{cases}$$

**Proof of Lemma 4.15.** Writing out the definitions we see that

$$\begin{aligned} \hat\beta - \beta &= (n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j-\overline{x})\,(\alpha+\beta x_j+E_j) - \beta \\ &= \alpha\,(n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j-\overline{x}) + \beta\,(n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j-\overline{x})\,x_j \\ &\qquad + (n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j-\overline{x})\,E_j - \beta \\ &= n^{-1}\,\rho_2^{-1/2}\,\sum_{j=1}^n w_j E_j, \end{aligned}$$

since $\sum_{j=1}^n (x_j-\overline{x}) = 0$ and $\sum_{j=1}^n (x_j-\overline{x})\,x_j = \sum_{j=1}^n (x_j-\overline{x})^2 = n\,\rho_2$. Moreover,

$$\begin{aligned} \hat\alpha - \alpha &= \overline{Y} - \alpha - \hat\beta\,\overline{x} \;=\; \alpha + \beta\,\overline{x} + \overline{E} - \alpha - \hat\beta\,\overline{x} \;=\; \overline{E} - (\hat\beta-\beta)\,\overline{x} \\ &= n^{-1}\,\sum_{j=1}^n E_j - n^{-1}\,\rho_2^{-1}\,\overline{x}\,\sum_{j=1}^n (x_j-\overline{x})\,E_j \\ &= n^{-1}\,\sum_{j=1}^n \rho_2^{-1/2}\,(\rho_2+\overline{x}^2)^{1/2}\,v_j\,E_j. \end{aligned}$$

Using that $\hat\alpha - \alpha = \overline{E} - (\hat\beta-\beta)\,\overline{x}$ and

$$\sum_{j=1}^n (x_j-\overline{x})\,(E_j-\overline{E}) \;=\; \sum_{j=1}^n (x_j-\overline{x})\,E_j \;=\; n\rho_2\,(\hat\beta-\beta),$$

we finally see that

$$\begin{aligned} \text{SSE} &= \sum_{j=1}^n (\alpha+\beta x_j+E_j-\hat\alpha-\hat\beta\,x_j)^2 \\ &= \sum_{j=1}^n ((\hat\beta-\beta)\,\overline{x} - \overline{E} + E_j - (\hat\beta-\beta)\,x_j)^2 \\ &= \sum_{j=1}^n ((E_j-\overline{E}) - (x_j-\overline{x})\,(\hat\beta-\beta))^2 \\ &= \sum_{j=1}^n (E_j-\overline{E})^2 - 2\,(\hat\beta-\beta)\,\sum_{j=1}^n (x_j-\overline{x})\,(E_j-\overline{E}) \\ &\qquad + (\hat\beta-\beta)^2\,\sum_{j=1}^n (x_j-\overline{x})^2 \\ &= \sum_{j=1}^n (E_j-\overline{E})^2 - 2\,n\,\rho_2\,(\hat\beta-\beta)^2 + n\,\rho_2\,(\hat\beta-\beta)^2 \\ &= \sum_{j=1}^n E_j^2 - n\,\overline{E}^2 - n\,\rho_2\,(\hat\beta-\beta)^2. \end{aligned}$$

Using the foregoing the statement easily follows. □

We may now express $\hat{\alpha}$, $\hat{\beta}$ and SSE in terms of $E$, and therefore the $\theta_k$ as well. In fact, writing

$$c_{j1} \;:=\; v_j, \quad c_{j2} \;:=\; w_j,$$

for $1 \le j \le n$, for both $k = 1, 2$ we have:

$$\theta_k \;=\; \theta_k(E) \;=\; (\textstyle\sum_{j=1}^n c_{jk} E_j)/\sqrt{\mathrm{SSE}(E)}.$$

Hence, the $\theta_k$ are of the form (4.34), with the $X_j$ replaced by $E_j$, taking

$$R(E) \;:=\; -\big(n\,\overline{E}^2 + n^{-1}\,(\textstyle\sum_{j=1}^n w_j E_j)^2\big).$$

For both $\theta_k$ we are dealing with an i.i.d. sample, the nuisance term $R(E)$ satisfying (4.33). Now let $q^2$ be the $q_0^2(\alpha, k)$ which we obtain from Lemma 4.11 by taking $\alpha = E_1$ and $k = n$, and let $M_j^2 := q^2\, n$ for all $j$. As in (4.96) then $s^2 = n\, q^2$, so that $s^2 \ne 0$ and $M_j^2 \le s^2$, for all $j$. Moreover:

**Lemma 4.16.** *For $k = 1, 2$,*

$$\textstyle\sum_{j=1}^n c_{jk}^2 = n \quad \textit{and} \quad \sum_{j=1}^n |c_{jk}|^3 \le 8n\,\rho_3\,\rho_2^{-3/2}.$$

**Proof of Lemma 4.16.** Obviously $\sum_{j=1}^n w_j^2 = \rho_2^{-1}\, n\,\rho_2 = n$, and furthermore $\sum_{j=1}^n |w_j|^3 = \rho_2^{-3/2}\, n\,\rho_3$. Moreover,

$$\begin{aligned}\rho_2^{-1}\,(\rho_2 + \overline{x}^2)\,\textstyle\sum_{j=1}^n v_j^2 \;&=\; n - 2\rho_2^{-1}\,\overline{x}\,\textstyle\sum_{j=1}^n (x_j - \overline{x}) + \rho_2^{-2}\,\overline{x}^2\, n\rho_2 \\ &=\; n\,\rho_2^{-1}\,(\rho_2 + \overline{x}^2),\end{aligned}$$

so that $\sum_{j=1}^n v_j^2 = n$, and, using the $C_r$-inequality and the inequality $|a|^{3/2} + |b|^{3/2} \le 2^{1/2}\,(|a| + |b|)^{3/2}$,

$$\begin{aligned}\rho_2^{-3/2}\,(\rho_2 + \overline{x}^2)^{3/2}\,\textstyle\sum_{j=1}^n |v_j|^3 \;&\le\; \textstyle\sum_{j=1}^n (1 + \rho_2^{-1}\,|\overline{x}|\,|x_j - \overline{x}|)^3 \\ &\le\; 4\,(n + \rho_2^{-3}\,|\overline{x}|^3\, n\,\rho_3) \;=\; 4n\,\rho_2^{-3/2}\,(\rho_2^{3/2} + \rho_3\,\rho_2^{-3/2}\,|\overline{x}|^3) \\ &\le\; 4n\,\rho_2^{-3/2}\,(\rho_3\,\rho_2^{-3/2})\,(\rho_2^{3/2} + |\overline{x}|^3) \\ &\le\; 8n\,(\rho_3\,\rho_2^{-3/2})\,\rho_2^{-3/2}\,(\rho_2 + \overline{x}^2)^{3/2},\end{aligned}$$

so that $\sum_{j=1}^{n} |v_j|^3 \le 8n\, \rho_3\, \rho_2^{-3/2}$. This finishes the proof. $\square$

The conditions of Lemma 4.6 are satisfied, which, together with the fact that in general $|a+b|^{3/2} \ll |a|^{3/2} + |b|^{3/2}$, leads to the fact that

$$\begin{aligned} \mathbb{E}\,|R(U)|^{3/2} &\ll \mathbb{E}\,|n\,\overline{U}^2|^{3/2} + \mathbb{E}\,|n^{-1}\,(\textstyle\sum_{j=1}^{n} w_j U_j)^2|^{3/2} \\ &= n^{-3/2}\,(\mathbb{E}\,|\textstyle\sum_{j=1}^{n} U_j|^3 + \mathbb{E}\,|\textstyle\sum_{j=1}^{n} w_j U_j|^3) \ \ll\ n^{-3/2} \ \le\ A_3^3, \end{aligned}$$

see as well (4.38). Now we are ready to apply Theorems 4.5 and 4.10 for $\theta_1$ and $\theta_2$. As to $A_1$, $A_2$, $A_3$ and $A_4$ we have the exact same situation as encountered in the proof of Corollary 4.12 (with the $X_j$ replaced by $E_j$), so that

$$\max\{A_1,\, A_2,\, A_3\} \ \ll\ B \quad \text{and} \quad \max\{A_1,\, A_2,\, A_3,\, A_4^{1/(1+\delta)}\} \ \ll\ C.$$

Lemma 4.16 is telling us that, for $k = 1, 2$,

$$\nu_{3k} \ =\ n^{-1} \textstyle\sum_{j=1}^{n} |c_{jk}|^3 \ \ll\ \rho_3\, \rho_2^{-3/2},$$

and the proof of (4.105) and (4.106) is easily concluded.

The statements (4.107) and (4.108) are easy consequences of (4.105) and (4.106), see the proof of Corollary 4.12. This finishes the proof of Corollary 4.14. $\square$

# Appendix A

# The Hoeffding decomposition

Let $X_1 : \Omega \to \mathcal{X}_1, \ \dots \ , X_n : \Omega \to \mathcal{X}_n$ be independent random variables, taking their values in arbitrary measurable spaces $(\mathcal{X}_j, \mathcal{B}_j)$. By $\sigma(A)$ we denote the $\sigma$-algebra generated by the $X_j$ with $j \in A$, that is to say, the smallest $\sigma$-algebra on $\Omega$ such that all $X_j$, $j \in A$ are measurable. We look at an arbitrary statistic

$$\mathbb{T} \ = \ \mathbb{T}(X_1, \dots , X_n), \tag{A.1}$$

where $\mathbb{T} : \prod_{j=1}^n \mathcal{X}_j \to \mathbb{R}$ is a measurable function, such that $\mathbb{E}\,|\mathbb{T}| < \infty$.

We describe a canonical decomposition for $\mathbb{T}$, due to Hoeffding (1948). Here we follow the lines of Van Zwet (1984) and Bentkus, Götze, Van Zwet (1997), Section 4, introducing as well the complementary concept of differences. We establish some basic facts and useful moment inequalities (see as well Lemmas 2.5 and 2.8).

Let

$$N \ := \ \{1, \dots , n\}.$$

Moreover, for any subset $A \subset N$, let $|A|$ denote its cardinality, and $A^c$ its complement. For any subset $A = \{j_1, \dots , j_k\} \subset N$, we now have the conditional expectation

$$\mathbb{E}\,(\mathbb{T}\,|\,A) \ = \ \mathbb{E}\,(\mathbb{T}\,|\,j_1, \dots , j_k) \ := \ \mathbb{E}\,(\mathbb{T}\,|\,X_{j_1}, \dots , X_{j_k}), \tag{A.2}$$

being the almost surely unique $\sigma(A)$-measurable random variable $S$ for which $\mathbb{E}\,|S| < \infty$ and $\mathbb{E}\,I\{G\}\,\mathbb{T} = \mathbb{E}\,I\{G\}\,S$ for all $G \in \sigma(A)$. By $I\{G\}$ we mean the indicator of the event $G$. Note that

$$\mathbb{E}(\mathbb{T}\,|\,\emptyset) = \mathbb{E}\,\mathbb{T} \qquad \text{and} \qquad \mathbb{E}(\mathbb{T}\,|\,N) = \mathbb{T}.$$

In case $\mathbb{E}\,\mathbb{T}^2 < \infty$, the mapping $\mathbb{T} \mapsto \mathbb{E}(\mathbb{T}\,|\,A)$ is the orthogonal projection on the Hilbert space $L^2(A)$ of equivalence classes of almost surely equal, square integrable functions $\varphi(X_j : j \in A)$ (using the standard inner product $\langle S_1, S_2\rangle := \mathbb{E}\,S_1 S_2$). We let

$$\mathbb{E}_A\mathbb{T} \;:=\; \mathbb{E}(\mathbb{T}\,|\,A^c). \tag{A.3}$$

For any $A \subset N$, we set

$$T_A \;=\; T_A(X_j : j \in A) \;:=\; \sum_{B\subset A}(-1)^{|A|-|B|}\,\mathbb{E}(\mathbb{T}\,|\,B). \tag{A.4}$$

As such, we interpret $T_A$ as a function of the $X_j$, $j \in A$, depending only on the distribution of the sample. For any set $A = \{j_1, \ldots, j_l\}$, we shall denote $T_A$ as well by $T_{j_1,\ldots,j_l}$. In this way we have that $T_\emptyset = \mathbb{E}\,\mathbb{T}$, and, for any $1 \le j \le n$ and $1 \le k < l \le n$,

$$T_j = \mathbb{E}(\mathbb{T}\,|\,j) - \mathbb{E}\,\mathbb{T}, \quad T_{k,l} = \mathbb{E}(\mathbb{T}\,|\,k,l) - \mathbb{E}(\mathbb{T}\,|\,k) - \mathbb{E}(\mathbb{T}\,|\,l) + \mathbb{E}\,\mathbb{T},$$

and so on. It is easily seen that for any two statistics $R$ and $S$ with existing first moments and $A \subset N$,

$$(R+S)_A \;=\; R_A + S_A. \tag{A.5}$$

We have the following inverse statement to (A.4):

**Lemma A.1.** *For all $B \subset N$,*

$$\mathbb{E}(\mathbb{T}\,|\,B) \;=\; \textstyle\sum_{C\subset B} T_C.$$

**Proof of Lemma A.1.** It is clear from (A.4) that $\sum_{C\subset B} T_C$ is a linear combination of $\mathbb{E}(\mathbb{T}\,|\,A)$ with $A \subset B$. Taking any such $A \subset B$, we see that it is encountered as a summand in the definition of $T_C$ precisely if $A \subset C$, and then with the factor $(-1)^{|C|-|A|}$. In this way, using Newton's binomial theorem, we see that the term $\mathbb{E}(\mathbb{T}\,|\,A)$ comes up

$$\textstyle\sum_{j=0}^{|B|-|A|}\binom{|B|-|A|}{j}(-1)^j \;=\; (-1+1)^{|B|-|A|} \;=\; I\{A = B\}$$

times (there being $\binom{|B|-|A|}{j}$ different choices of $C$ for which $A \subset C \subset B$ and $|C| - |A| = j$). Thus we see that indeed $\sum_{C \subset B} T_C = \mathbb{E}(\mathbb{T} \mid B)$. $\square$

In particular we have that

$$\mathbb{T} \;=\; \textstyle\sum_{C \subset N} T_C \;=\; \mathbb{E}\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2 + \mathbf{T}_3, \tag{A.6}$$

taking

$$\mathbb{T}_1 \;:=\; \textstyle\sum_{j=1}^n T_j, \quad \mathbb{T}_2 \;:=\; \textstyle\sum_{1 \le j < k \le n} T_{j,k}, \quad \mathbf{T}_3 \;:=\; \textstyle\sum_{A:\,|A| \ge 3} T_A, \tag{A.7}$$

which states the so-called *Hoeffding decomposition* of $\mathbb{T}$. Here $\mathbb{T}_1$ is the so-called *linear part* of $\mathbb{T}$. Note that $T_A$ is in general a random variable depending on the $X_j$ with $j \in A$. The crucial feature of the Hoeffding decomposition is that it is *orthogonal* in $L^2(N)$, that is to say, in case $\mathbb{E}\mathbb{T}^2 < \infty$ we have that $\mathbb{E}T_A T_B = 0$ for all $A \neq B$.

Complementary to the Hoeffding decomposition is the concept of so-called *differences*. To this, for any statistic $S = S(X_1, \ldots, X_n)$ with $\mathbb{E}|S| < \infty$ and any $j \in N$ we define

$$D_j S \;:=\; S - \mathbb{E}(S \mid \{X_k \,:\, k \neq j\}) \;=\; S - \mathbb{E}_j S. \tag{A.8}$$

In this way always $D_j\, D_j S = D_j S$ and $D_l\, D_m S = D_m\, D_l S$, so that one may also speak of

$$D_A S \;:=\; D_{j_1} \cdots D_{j_k} S$$

for any non-empty set $A = \{j_1, \ldots, j_k\}$. Note that $D_j$ is linear in the sense that always

$$D_j\, \lambda\, S = \lambda\, D_j S \quad \text{and} \quad D_j\,(S_1 + S_2) = D_j S_1 + D_j S_2, \tag{A.9}$$

and that for any statistic $S_1 = S_1(X_k \,:\, k \neq j)$ which does not depend on $X_j$,

$$D_j S_1 S_2 \;=\; S_1\, D_j S_2. \tag{A.10}$$

Returning to $\mathbb{T}$, we have the following general equalities:

**Lemma A.2.** *For any non-empty set $A \subset N$,*

$$D_A \mathbb{T} \;=\; \textstyle\sum_{B \subset A} (-1)^{|B|}\, \mathbb{E}(\mathbb{T} \mid B^c) \quad \textit{and} \quad T_A \;=\; \mathbb{E}(D_A \mathbb{T} \mid A). \tag{A.11}$$

*Moreover, for any non-empty set $A \subset N$,*

$$D_A \mathbb{T} \;=\; \textstyle\sum_{B:\, A \subset B} T_B. \tag{A.12}$$

**Proof of Lemma A.2.** A proof of the first equality is easily found using induction on the number of elements of $A$. The statement is obvious in the case that $|A| = 1$. Now suppose the statement is correct for all $A$ with $|A| = k$, for some $1 \le k \le n-1$. Looking at any set $A \cup \{m\}$ of size $k+1$, with $m \notin A$, using the inductional hypothesis we see that

$$D_{A\cup\{m\}}\mathbb{T} \;=\; D_m\, D_A \mathbb{T} \;=\; \textstyle\sum_{B\subset A} (-1)^{|B|}\, D_m \mathbb{E}\,(\mathbb{T}\,|\,B^c).$$

Here $D_m \mathbb{E}\,(\mathbb{T}\,|\,B^c) = \mathbb{E}\,(\mathbb{T}\,|\,B^c) - \mathbb{E}\,(\mathbb{T}\,|\,(B \cup \{m\})^c)$, so that

$$\begin{aligned} D_{A\cup\{m\}}\mathbb{T} \;&=\; \textstyle\sum_{C\subset A\cup\{m\},\, m\notin C} (-1)^{|C|}\, \mathbb{E}\,(\mathbb{T}\,|\,C^c) \\ &\qquad\quad + \textstyle\sum_{C\subset A\cup\{m\},\, m\in C} (-1)^{|C|}\, \mathbb{E}\,(\mathbb{T}\,|\,C^c) \\ &=\; \textstyle\sum_{C\subset A\cup\{m\}} (-1)^{|C|}\, \mathbb{E}\,(\mathbb{T}\,|\,C^c), \end{aligned}$$

which proves the point.

The second equality is a consequence of the first, as for any $B \subset A$ we have that $\mathbb{E}\,(\mathbb{E}\,(\mathbb{T}\,|\,B^c)\,|\,A) = \mathbb{E}\,(\mathbb{T}\,|\,A \cap B^c)$, so that, using the substitution $C := A \cap B^c$ and (A.4),

$$\begin{aligned} \mathbb{E}\,(D_A \mathbb{T}\,|\,A) \;&=\; \textstyle\sum_{B\subset A} (-1)^{|B|}\, \mathbb{E}\,(\mathbb{T}\,|\,A\cap B^c) \\ &=\; \textstyle\sum_{C\subset A} (-1)^{|A|-|C|}\, \mathbb{E}\,(\mathbb{T}\,|\,C) \;=\; T_A. \end{aligned}$$

We turn to (A.12). In case $A = \{j\}$, from Lemma A.1 it follows that

$$\begin{aligned} D_j \mathbb{T} \;&=\; \mathbb{T} - \mathbb{E}\,(\mathbb{T}\,|\,N\backslash\{j\}) \\ &=\; \textstyle\sum_B T_B - \sum_{B:\, j\notin B} T_B \;=\; \sum_{B:\, j\in B} T_B, \end{aligned}$$

which is what we need in the case where $|A| = 1$. Now suppose that the statement is correct for all $A$ with $|A| = k$, for some $1 \le k \le n-1$. Looking at any set $A \cup \{m\}$ of size $k+1$, with $m \notin A$, we see that

$$D_{A\cup\{m\}}\mathbb{T} \;=\; D_m\, D_A \mathbb{T} \;=\; \textstyle\sum_{B:\, A\subset B} D_m T_B \;=\; \sum_{B:\, A\cup\{m\}\subset B} T_B,$$

using the fact that $D_m T_B = I\{m \in B\}\, T_B$. As to the latter: in case $m \notin B$, using (A.10) we see that $D_m T_B = T_B\, D_m 1 = 0$, whereas otherwise $D_m T_B = T_B - \mathbb{E}_m T_B = T_B$. This finishes the proof. $\square$

The Hoeffding decomposition (A.6) and the moments $\mathbb{E}\,|T_A|^l$ of its parts are natural tools for the mathematical analysis of the statistic $\mathbb{T}$. However, in applications, the differences $D_A\mathbb{T}$ and their moments $\mathbb{E}\,|D_A\mathbb{T}|^l$ are often more convenient to calculate or estimate. To this: using (A.11) and Jensen's inequality (conditional version), for any non-empty set $A \subset N$ and $l \geq 1$ we have:

$$\mathbb{E}\,|T_A|^l \;=\; \mathbb{E}\,|\mathbb{E}\,(D_A\mathbb{T}\,|\,A)|^l \;\leq\; \mathbb{E}\mathbb{E}\,(|D_A\mathbb{T}|^l\,|\,A) \;=\; \mathbb{E}\,|D_A\mathbb{T}|^l. \tag{A.13}$$

In the non-identically distributed case we will make use of the following moments corresponding to $\mathbb{T}$:

$$\sigma^2 := \operatorname{var}\mathbb{T}, \quad s_j^2 := \mathbb{E}\,T_j^2, \quad s^2 := \operatorname{var}\mathbb{T}_1 = \textstyle\sum_{j=1}^n s_j^2, \tag{A.14}$$

$$\beta_j := \mathbb{E}\,|T_j|^3, \quad \beta := \textstyle\sum_{j=1}^n \beta_j, \quad \Delta^2 \;:=\; \textstyle\sum_{1\leq j<k\leq n} \mathbb{E}\,|D_j D_k \mathbb{T}|^2. \tag{A.15}$$

We will denote as well, see (A.12),

$$\Delta^2 \;=\; \textstyle\sum_{1\leq j<k\leq n} \mathbb{E}\,T_{j,k}^2 + \sum_{1\leq j<k\leq n}\, \sum_{j,k\in A,\,|A|\geq 3} \mathbb{E}\,T_A^2 \;=:\; \Delta_1^2 + \Delta_2^2. \tag{A.16}$$

Note that

$$\Delta_1^2 = \operatorname{var}\mathbb{T}_2 \quad \text{and} \quad \Delta_2^2 \geq \operatorname{var}\mathbf{T}_3. \tag{A.17}$$

Obviously (because of orthogonality) we have in general that

$$\sigma^2 \;=\; s^2 + \operatorname{var}\mathbb{T}_2 + \operatorname{var}\mathbf{T}_3,$$

so that for example $s^2 \leq \sigma^2$.

Looking at (A.16), we notice that for any fixed subset $C$ of $N$ with $|C| = l$, $l \geq 2$, the term $\mathbb{E}\,T_C^2$ is encountered $\binom{l}{2}$ times (as this is the number of possible choices for the pair $(j,k)$). As a result, for all $1 \leq l \leq n$ taking

$$\mathbb{T}_l \;:=\; \textstyle\sum_{A:\,|A|=l} T_A,$$

we have that

$$\Delta^2 \;=\; \sum_{l=2}^{n} \binom{l}{2} \sum_{C:\,|C|=l} \mathbb{E} T_C^2 \;=\; \sum_{l=2}^{n} \binom{l}{2} \operatorname{var} \mathbb{T}_l. \tag{A.18}$$

A bit more precise analysis tells us that

$$\Delta_1^2 \;=\; \sum_{C:\,|C|=2} \mathbb{E} T_C^2 \quad \text{and} \quad \Delta_2^2 \;=\; \sum_{l=3}^{n} \binom{l}{2} \sum_{C:\,|C|=l} \mathbb{E} T_C^2. \tag{A.19}$$

In case of an i.i.d. sample and a statistic that is symmetric in its arguments, the following moment expressions are essential:

$$\tilde\beta_3 := \mathbb{E}|n^{1/2}\, T_1|^3, \quad \tilde\gamma_p := \mathbb{E}|n^{3/2}\, T_{1,2}|^p, \quad \tilde\Delta_m^p := \mathbb{E}|n^{m-1/2}\, D_1 \cdots D_m \mathbb{T}|^p, \tag{A.20}$$

with $m \in N$ and $p \geq 0$, and

$$\tilde\delta_2 \;:=\; \tilde\Delta_2^2 - \tilde\gamma_2 \;=\; n^3 \sum_{A:\,\{1,2\}\subset A,\,|A|\geq 3} \mathbb{E} T_A^2, \tag{A.21}$$

cf. (A.12). Being more precise, considering the fact that there are $\binom{n-2}{l-2}$ possible sets $A$ such that $1,2 \in A$ and $|A| = l$, we see that

$$\tilde\gamma_2\, n^{-1} = n^2\, \mathbb{E} T_{1,2}^2 \quad \text{and} \quad \tilde\delta_2\, n^{-1} = n^2 \sum_{l=3}^{n} \binom{n-2}{l-2} \mathbb{E} T_{1,\ldots,l}^2, \tag{A.22}$$

and, of course,

$$\tilde\Delta_2^2\, n^{-1} \;=\; n^2 \sum_{l=2}^{n} \binom{n-2}{l-2} \mathbb{E} T_{1,\ldots,l}^2, \tag{A.23}$$

which will be one of the crucial moment expressions in our i.i.d. bounds. The following relation exists between $\Delta^2$ and $\tilde\Delta_2^2$:

$$\Delta^2 \;=\; \binom{n}{2} \sum_{\{1,2\}\subset A} \mathbb{E} T_A^2 \;=\; \tfrac{1}{2}(1-n^{-1})\; \tilde\Delta_2^2\, n^{-1} \;\approx\; \tfrac{1}{2}\, \tilde\Delta_2^2\, n^{-1}. \tag{A.24}$$

More precise, we have the following:

$$\Delta_1^2 \;=\; \tfrac{1}{2}(1-n^{-1})\; \tilde\gamma_2\, n^{-1} \;\approx\; \tfrac{1}{2}\, \tilde\gamma_2\, n^{-1} \tag{A.25}$$

and

$$\Delta_2^2 \;=\; \tfrac{1}{2}(1-n^{-1})\; \tilde\delta_2\, n^{-1} \;\approx\; \tfrac{1}{2}\, \tilde\delta_2\, n^{-1}. \tag{A.26}$$

# Bibliography


[1] I.B. Alberink, A Berry-Esseen bound for $U$-statistics in the non-i.i.d. case, Preprint 98-057 SFB 343, University of Bielefeld, 1998.

[2] I.B. Alberink, V. Bentkus, Bounds for the concentration of asymptotically normal statistics, 1999, Acta Applic. Math., ????, p.??.

[3] I.B. Alberink, W.R. Pestman, *Mathematical statistics. Problems and detailed solutions*, Walter De Gruyter, 1998, Berlin.

[4] A.K. Aleshkyavichene, V.A. Statulyavichus, Inversion formulas in the case of a discontinuous limit law, 1997, *Theory Prob. Appl. 42, p.1-16.*

[5] V. Bentkus, On the asymptotical behavior of the constant in the Berry-Esseen inequality, 1994, *J. Theoret. Probab. 7, p.211-224.*

[6] V. Bentkus, M. Bloznelis, F. Götze, A Berry-Esséen Bound for Student's Statistic in the Non-I.I.D. Case, 1996, *Journal of Theoretical Probability 9, p.765-796.*

[7] V. Bentkus, F. Götze, The Berry-Esseen bound for Student's statistic, 1996, *Ann. Prob. 24, p.491-503.*

[8] V. Bentkus, F. Götze, Optimal results of convergence in the CLT for quadratic forms, 1996, *Ann. Prob. 24, p.??.*

[9] V. Bentkus, F. Götze and W.R. van Zwet, An Edgeworth expansion for symmetric statistics, 1997, *Ann. Statist. 25, p.851-896.*

[10] G. Blom, Some properties of incomplete $U$-statistics, 1976, *Biometrika 63, p.573-580.*

[11] M. Bloznelis, A. Rackauskas, A Berry-Esséen bound for least squares estimators of regression parameters, 1999, *Lithuan. Math. J. 39, p.1-7.*

[12] H. Callaert, P. Janssen, The Berry-Esseen theorem for $U$-statistics, 1978, *Ann. Statist. 6, p.417-421.*

[13] Y.-K. Chan, J. Wierman, On the Berry-Esseen theorem for $U$-statistics, 1977, *Ann. Prob. 5, p.136-139.*

[14] R.J.M.M. Does, Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis, 1982, *Ann. Prob. 10, p.982-991.*

[15] E.B. Dynkin, A. Mandelbaum, Symmetric statistics, Poisson point processes, and multiple Wiener integrals, 1983, *Ann. Statist. 11, p.739-745.*

[16] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd Ed., Wiley, 1971, New York.

[17] K.O. Friedrich, A Berry-Esseen bound for functions of independent random variables, 1989, *Ann. Statist. 17, p.170-183.*

[18] J.D. Gibbons, S. Chakraborti, *Nonparametric statistical inference*, 3rd Ed., Marcel Dekker, Inc., 1992, New York.

[19] L. Giraitis, D. Surgailis, A limit theorem for a triangular array of symmetric statistics, 1984, *Statistics and control of stochastic processes, Moscow, p.147-166.* Transl. Ser. Math. Engrg., Optim. Software, New York, 1985.

[20] Z. Guan, A Berry-Esseen bound for $k$-sample symmetric statistics, 1994, *Northeast. Math. J. 10, p.411-420.*

[21] R. Helmers, A Berry-Esseen theorem for linear combinations of order statistics, 1981, *Ann. Prob. 9, p.342-347.*

[22] R. Helmers, W.R. van Zwet, The Berry-Esseen bound for $U$-statistics, *Statistical Decision Theory and Related Topics III*, Vol. 1 (S.S. Gupta and J.O. Berger, eds.), p.497-512, Academic Press, 1982, New York.

[23] W. Hengartner, R. Theodorescu, *Concentration functions*, Academic Press, 1973, New York.

[24] W. Hoeffding, A class of statistics with asymptotically normal distributions, 1948, *Ann. Math. Stat. 19, p.293-325.*

[25] S. Janson, The asymptotic distributions of incomplete $U$-statistics, 1984, *Z. Wahrsch. verw. Gebiete 66, p.495-505.*

[26] T.L. Lai, J.Q. Wang, Edgeworth expansions for symmetric statistics with applications to bootstrap methods, 1993, *Statist. Sinica 3, p.517-542.*

[27] M. Loève, *Probability theory I*, Springer, 1977, New York.

[28] A. Mandelbaum, M.S. Taqqu, Invariance principle for symmetric statistics, 1984, *Ann. Statist. 12, p.483-496.*

[29] V. Mandrekar, M.M. Meerschaert, Sample moments and symmetric statistics, 1994, *Stochastic analysis on infinite-dimensional spaces* (Baton Rouge, L.A.), p.197-210, Pitman Res. Notes Math. Ser., 310, Longman Sci. Tech., Harlow.

[30] W.R. Pestman, *Mathematical Statistics, An Introduction*, Walter de Gruyter, 1998, Berlin.

[31] V.V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Clarendon Press, 1995, Oxford.

[32] H. Prawitz, Limits for a distribution, if the characteristic function is given in a finite doamin, 1972, *Skand. Aktuar. Tidskr., p.138-154.*

[33] H. Putter, W.R. van Zwet, Empirical Edgeworth expansions for symmetric statistics, 1998, *Ann. Statist. 26, p.1540-1569.*

[34] Q. Wang, $L_p$ rates of convergence for symmetric statistics, 1998, *Nanjing Daxue Xuebao Shuxue Bannian Kan 15, p.41-49.*

[35] W.R. van Zwet, A Berry-Esseen bound for symmetric statistics, 1984, *Z. Wahrsch. verw. Gebiete 66, p.425-440.*

# Samenvatting

Zij $X_1, \ldots, X_n$ een steekproef van onafhankelijke en identiek verdeelde, reëelwaardige stochasten, zodanig dat $\mathbb{E}\,X_1^2 > 0$ en $\mathbb{E}\,|X_1|^3 < \infty$. Neem voor het gemak aan dat $\mathbb{E}\,X_1 = 0$, en zij $\sigma^2 := \mathbb{E}\,X_1^2$. Het is een bekend resultaat dat het genormaliseerde steekproefgemiddelde, dat wil zeggen, de steekproefgrootheid

$$\mathbb{T} \;:=\; n^{-1/2}\,\sigma^{-1}\,(X_1 + \ldots + X_n),$$

in verdeling convergeert naar de standaard-normale verdeling. Een sterker resultaat is het feit dat, als $\Phi$ staat voor de verdelingsfunctie horend bij de standaard-normale verdeling,

$$D(\mathbb{T}) \;:=\; \sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T} \le x) - \Phi(x)| \;\le\; 3n^{-1/2}\,\sigma^{-3}\,\mathbb{E}\,|X_1|^3.$$

Dit type bovengrens, dat de convergentie-snelheid beschrijft van een steekproefgrootheid $\mathbb{T}$ naar een normale verdeling, wordt gewoonlijk aangeduid met de naam Berry-Esseen grens, en bestaat ook voor meer algemene $\mathbb{T}$.

Een verwant begrip is dat van de concentratie-grens. Voor een willekeurige grootheid $\mathbb{T}$ die een functie is van de steekproef wordt diens concentratiefunctie gedefinieerd door

$$Q(\mathbb{T},\lambda) \;:=\; \sup_{x\in\mathbb{R}} \mathbb{P}(x \le \mathbb{T} \le x+\lambda) \quad (\lambda \ge 0),$$

zodat deze staat voor de concentratie van kans-massa op (kleine) intervallen. In Hoofdstuk 2 van dit proefschrift worden bovengrenzen afgeleid voor $Q(\mathbb{T},\lambda)$, uitgaande van stochasten die onafhankelijk maar *niet* noodzakelijk identiek verdeeld zijn. Om dit voor elkaar te krijgen wordt $\mathbb{T}$ beschouwd door middel van haar Hoeffding decompositie

$$\mathbb{T} \;=\; \mathbb{E}\,\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2 + \mathbf{T}_3,$$

zie Appendix A. Het lineaire deel $\mathbb{T}_1$ is de beste benadering van $\mathbb{T} - \mathbb{E}\,\mathbb{T}$ van de vorm $\sum_{j=1}^n \varphi_j(X_j)$, en $\mathbb{T}_1 + \mathbb{T}_2$ is de beste benadering door middel van een som over functies horende bij elk paar observaties $(X_j, X_k)$ afzonderlijk. De variantie van het lineaire deel aangevend met $s^2$, leiden we een concentratie-grens voor $\mathbb{T}/s$ af in termen van een maximum over $\lambda$, de som over de absolute derde momenten van de sommanden van $\mathbb{T}_1$, en een uitdrukking $\Delta^2$ die sterk lijkt op de variantie van het niet-lineaire deel $\mathbb{T}_2 + \mathbf{T}_3$. Het verkregen resultaat lijkt optimaal.

In Hoofdstuk 3 worden Berry-Esseen grenzen afgeleid voor willekeurige steekproefgrootheden $\mathbb{T}$. De daadwerkelijke bovengrens is van dezelfde vorm als de concentratie-grens, maar zonder de term $\lambda$, en met $(\Delta^2)^{1/2}$ in plaats van $\Delta^2$. In plaats van $\Phi$ wordt ook wel een korte Edgeworth-ontwikkeling $G = \Phi + \Phi_1$ in de definitie van $D(\mathbb{T})$ gebruikt, wat leidt tot een resultaat met een hogere (=betere) macht van $\Delta^2$. Een belangrijk aspect van het bewijs is een randomisatie van de steekproef-elementen, die het mogelijk maakt met niet-identiek verdeelde situaties om te gaan als met identiek verdeelde.

In Hoofdstuk 4 worden toepassingen gegeven van de belangrijkste resultaten. Achtereenvolgens worden eenvoudige lineaire rang-grootheden, $U$-statistics, incomplete $U$-statistics en zelf-genormaliseerde steekproefgrootheden bekeken, waarvoor de juiste concentratie-grenzen worden afgeleid. De verkregen Berry-Esseen grenzen zijn over het algemeen optimaal, en soms, bijvoorbeeld voor zelf-genormaliseerde steekproefgrootheden, iets slechter dan de beste resultaten (in de zin dat we $(3+\delta)^{\mathrm{e}}$ momenten nodig hebben in plaats van $3^{\mathrm{e}}$).

# Summary

Consider a sample of independent, identically distributed, real-valued random variables $X_1, \ldots, X_n$, for which $\mathbb{E}\, X_1^2 > 0$ and $\mathbb{E}\,|X_1|^3 < \infty$. For convenience, assume that $\mathbb{E}\, X_1 = 0$ and write $\sigma^2 := \mathbb{E}\, X_1^2$. Then it is well-known that the normalized sample mean, that is, the statistic

$$\mathbb{T} \;:=\; n^{-1/2}\, \sigma^{-1}\, (X_1 + \ldots + X_n),$$

converges in distribution to the standard normal distribution. A stronger result is the fact that, writing $\Phi$ for the standard normal distribution function,

$$D(\mathbb{T}) \;:=\; \sup_{x \in \mathbb{R}} |\mathbb{P}(\mathbb{T} \leq x) - \Phi(x)| \;\leq\; 3n^{-1/2}\, \sigma^{-3}\, \mathbb{E}\,|X_1|^3.$$

This type of bound, describing the rate of convergence of $\mathbb{T}$ to a normal distribution, is usually referred to as Berry-Esseen bound, and exists as well for more general $\mathbb{T}$.

A related concept is that of concentration bounds. For any statistic $\mathbb{T}$ which is a function of the sample, its concentration function is defined by

$$Q(\mathbb{T}, \lambda) \;:=\; \sup_{x \in \mathbb{R}} \mathbb{P}(x \leq \mathbb{T} \leq x + \lambda) \quad (\lambda \geq 0),$$

thus reflecting the concentration of probability mass on (small) intervals. In Chapter 2 of the thesis upper bounds are derived for $Q(\mathbb{T}, \lambda)$, starting from random variables which are independent but *not* necessarily identically distributed. To do this we consider $\mathbb{T}$ through its Hoeffding decomposition

$$\mathbb{T} \;=\; \mathbb{E}\,\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2 + \mathbf{T}_3,$$

see Appendix A. The linear part $\mathbb{T}_1$ is the best approximation of $\mathbb{T} - \mathbb{E}\,\mathbb{T}$ of the form $\sum_{j=1}^n \varphi_j(X_j)$, and $\mathbb{T}_1 + \mathbb{T}_2$ is its best approximation by a sum over

functions of each pair $(X_j, X_k)$ of observations. Writing $s^2$ for the variance of the linear part, a concentration bound is attained for $\mathbb{T}/s$ in terms of a maximum over $\lambda$, the sum over the absolute third moments of $\mathbb{T}_1$'s summands, and an expression $\Delta^2$ very similar to the variance of the non-linear part $\mathbb{T}_2 + \mathbf{T}_3$. In reality this is a rather satisfactory bound.

In Chapter 3 Berry-Esseen bounds are derived for general statistics $\mathbb{T}$, of the form just mentioned. The actual bound is of the same form as the concentration bound, though of course without the term $\lambda$, with $(\Delta^2)^{1/2}$ instead of $\Delta^2$. Instead of $\Phi$ we use as well a short Edgeworth expansion $G = \Phi + \Phi_1$ in the definition of $D(\mathbb{T})$, in which case a bound is obtained that has a higher (=better) power of $\Delta^2$. An important aspect of the proof is a randomization of the sample elements, which makes it possible to cope with non-i.i.d. situations as with i.i.d. ones.

In Chapter 4 applications are given of the main results. In turn simple linear rank statistics, $U$-statistics, incomplete $U$-statistics and self-normalized statistics are considered, for which the right concentration bounds are derived, and Berry-Esseen bounds that sometimes equal the best known results and sometimes, for example for self-normalized statistics, are a bit worse (requiring $3 + \delta^{\text{th}}$ moments instead of $3^{\text{rd}}$).

# Acknowledgements

At this place I would like to thank all the people who have, directly or indirectly, contributed to my dissertation.

Here I will start by saying a few words about the person who stood at the beginning of all of this, Wiebe Pestman. His excellent teaching, and constant encouragements to actually understand the mathematics I was dealing with, turned me into the honest, hard-working mathematician that I am today. It was also Wiebe who suggested to me to stay at the university in the first place, which I probably would not have considered otherwise. Wiebe's great teaching skills and profound knowledge of many fields of mathematics will be missed in Nijmegen.

Secondly I would like to thank my promotor Frank den Hollander, for logistic support, and my supervisor Martien van Zuijlen, who never got tired of carefully checking my manuscripts, technical detail by technical detail, nor of making suggestions about ways to improve exposition.

This thesis' turning out as it is at the moment is largely attributable to Vidas Bentkus, whom I visited a lot at the University of Bielefeld during the last two years. Vidas usually went through lots of trouble - and was rather successful in - persuading me not to go about wasting days doing lengthy (and useless) calculations. He showed me what the obtaining of Berry-Esseen bounds is really about, and gave me lots of good suggestions as to both obtainable results and methods in order to prove them. The bulk of Chapter 2 about concentration functions may be found in a joint paper by the two of us. Further results are mine, although all did arise from the numerous conversations that we had on the topic of Berry-Esseen bounds. It has always been a great pleasure working with and learning from him.

I would like to take the opportunity as well to thank Gyula Pap and Mindaugas Bloznelis, who regularly looked at my results and gave sound advice about it. Gyula I thank as well for having me over at the University of Debrecen for a month in July 1997. Mindaugas has been great company during the many long weeks spent in Bielefeld, enlightening me on such various subjects as Lithuanian history, Russian soups and dialectic materialism. Thanks as well to my landlady Ulrike Rehring, for going way beyond my knowledge of German and always providing me with a place to sleep.

Turning back to Nijmegen, I would like to thank my room mates Jan and Anita for excellent company. Jan for being the perfect sparring partner at table tennis, Anita for not minding about my being a constant source of distraction. Thanks to Martijn for Beatnik Filmstars, to Peer for Optiganally Yours and to Marjolein for the Young Fresh Fellows. Thanks to Jozsef for showing me around the Nijmegen pubs.

Finally I would like to thank my parents, for moral (and financial) support, and all friends and family that I did not mention by name, but who were important anyway.

# Curriculum Vitae

I was born in Hengelo (Ov), June 5, 1971. Here I attended the Lyceum de Grundel from 1983 up to 1989. After my graduation I went to study philosophy at the University of Nijmegen, the first-year examination of which I passed in 1990. In September 1990 I started studying mathematics, in which I graduated (cum laude) August 1995. My Master's thesis was on polynomial estimators with minimal variance, under supervision of Dr. W.R. Pestman. While not completing the philosophy study because of a lack of time, in September 1995 I started working as a Ph.D. student at the mathematics department of the University of Nijmegen. Here I worked in the field of statistics, under supervision of Dr. M.C.A. van Zuijlen and Prof. W.Th.F. den Hollander. During the last four years I have paid several visits to the University of Debrecen (Hungary) and the University of Bielefeld (Germany), where I have worked with Prof. Dr. G. Pap and Dr. V. Bentkus.